AD_____

Award Number: W81XWH-09-1-0169

TITLE:

Control of Disease Recurrence by Tumor-Infiltrating T Cells in Ovarian Cancer

PRINCIPAL INVESTIGATOR:

Brad H. Nelson, Ph.D.

CONTRACTING ORGANIZATION:

British Columbia Cancer Agency,

Victoria BC V8R 6V5
Canada

REPORT DATE:

March 2010

TYPE OF REPORT:

Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT:

X Approved for public release; distribution unlimited



`

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. REPORT DATE *(DD-MM-YYYY)* | 2. REPORT TYPE | 3. DATES COVERED *(From - To)* |
|---|---|---|
| 01-03-2010 | Annual Report | 1 Mar 2009 - 28 Feb 2010 |

**4. TITLE AND SUBTITLE**
Control of Disease Recurrence by Tumor-Infiltrating T Cells in Ovarian Cancer

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**
W81XWH-09-1-0169

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
Brad H. Nelson, Ph.D.

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
British Columbia Cancer Agency,

Victoria BC V8R 6V5
Canada

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
US Army Medical Research and Material Command
Fort Detrick, Maryland
21702-5012

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for public release; distribution unlimited

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT** Ovarian cancer patients with large numbers of T cells in their tumor live longer after chemotherapy compared to patients with fewer T cells in their tumor. Our goal is to use modern genomic techniques to identify the antigens recognized by these T cells, with an emphasis on new antigens that arise during chemotherapy. To this end, we are collecting matched primary and recurrent tumor tissue from ovarian cancer patients (Tasks 1-2). This is progressing well, but the precise timeline is beyond our control as it depends (most unfortunately) on patients suffering a recurrence. In the meantime, we are also testing blood samples from patients for the emergence of new antibody responses during chemotherapy (Task 3). Tumor tissue from one patient to date is being subjected to "whole transcriptome shotgun sequencing" (Task 4). Finally, new methods are being developed to rapidly test the recognition of tumor mutations by T cells from patients (Task 5). Overall, this project is progressing on schedule and is already yielding publishable results.

**15. SUBJECT TERMS**
Tumor immunology, immunotherapy, ovarian cancer, antibody, T cell, tumor antigen, mutation, next generation sequencing

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | USAMRMC |
| U | U | U | UU | 97 | 19b. TELEPHONE NUMBER *(include area code)* |

**Table of Contents**

**Page**

**W81XWH-09-1-0169 (OC080380) Annual Report, March 2010**

*PI:* Brad H. Nelson, Ph.D.
*Co-PIs:* Rob Holt, Ph.D., John Webb Ph.D., Peter Watson, M.D.

*Title of Project:* Control of Disease Recurrence by Tumor-Infiltrating T Cells in Ovarian Cancer

## INTRODUCTION:

In epithelial ovarian cancer (EOC), tumor-infiltrating CD8+ T cells are strongly associated with increased progression-free and overall survival following chemotherapy. However, the antigens recognized by tumor-infiltrating T cells remain largely unknown. Furthermore, it is not clear how tumor-infiltrating T cells, having failed to prevent the primary tumor, can oppose tumor recurrence. Recent work has shown that chemotherapy, in the context of impaired DNA repair pathways, induces mutations in the cancer genome, some of which contribute to chemo-resistance. Chemotherapy-induced mutations may provide a new source of tumor antigens for T cells, since mutated or aberrantly expressed proteins should be perceived as "non-self". Based on these considerations, we hypothesize that the mutational effects of chemotherapy generate new tumor antigens that trigger a second wave of CD8+ T cell responses, which in turn promote favorable clinical outcomes. To test this hypothesis, we are using serological and genomic methods to identify the evolving repertoire of tumor antigens and T cell responses in EOC patients who demonstrate favorable clinical outcomes.

The study has five tasks:

Task 1. Collection and processing of biospecimens

Task 2. To determine whether chemotherapy induces the emergence of new tumour-associated CD8+ T cell clones in EOC.

Task 3. To identify by serological approaches tumor antigens induced by chemotherapy in EOC.

Task 4. To identify changes to the tumor transcriptome induced by chemotherapy in EOC.

Task 5. To determine whether tumor-associated CD8+ T cells in EOC recognize putative chemotherapy-induced antigens.

*Significance.* This will be the first study to test the hypothesis that the mutational effects of platinum/taxane-based chemotherapy generate novel antigens that stimulate host CD8+ T cell responses. With a better understanding of how T cell responses evolve during standard treatments, we believe it will be possible in future to prolong progression-free survival by enhancing the host T cell response using immunodulatory agents, vaccines or adoptively transferred T cells.

## BODY:

### *Task 1. Collection and processing of biospecimens.*

This task concerns the collection of matched primary and recurrent ascites from 6 patients with high-grade serous ovarian cancer; isolation and storage of CD45+ leukocyte,

CD8+ lymphocyte and CD45- tumor cell subfractions; and collection of blood samples before, during and after primary surgery and chemotherapy.

***Progress to date:*** As anticipated in the original proposal, we are currently focusing on specimen collection. To date, we have collected 3 matched primary and recurrent ascites tumor specimens. However, the main challenge is to collect these from women with a favorable progression-free interval (PFI) (ideally > 24 months). Currently, we have 7 patients who have donated primary tumor specimens already and who are approaching our PFI criterion. We are making every effort to monitor these patients and collect recurrent ascites specimens as they become available.

***Task 2. To determine whether chemotherapy induces the emergence of new tumour-associated CD8+ T cell clones in EOC.***

We proposed to study tumor-infiltrating T cells from primary and recurrent ascites to look for changes in T cell subsets (CD3, CD4, CD8), and various activation and differentiation markers. We also proposed to sequence T cell receptors (TCRs) to identify 10-20 predominant T cell clones from recurrent tumors. Serial blood and ascites samples will then be analyzed by QPCR with clonotype-specific primers to determine the time of emergence of these T cell clones. Our hypothesis predicts that a large proportion of CD8+ T cell clones present in recurrent ascites will have arisen during or after chemotherapy.

***Progress to date:***

We have started experiments for Task 2 using specimens from a patient (IROC024) who fell a few months short of our PFI criterion, but is nonetheless of great interest for several reasons: (a) we have collected primary ascites as well as matched ascites from her first, second and third recurrence; (b) there are ample vials of tumor cells available from each time point, so we can use these specimens to develop our methodology without fear of squandering precious samples; and (c) a full complement of serial blood samples is available for later experiments to track T cell clones over time. Using these specimens, we have successfully separated tumor cells and T cells by flow cytometry, and these samples have now been sent to Rob Holt's lab for RNA isolation, whole transcriptome sequencing, and TCR sequencing. **Figure 1** summarizes the clinical history and available specimens for IROC024.

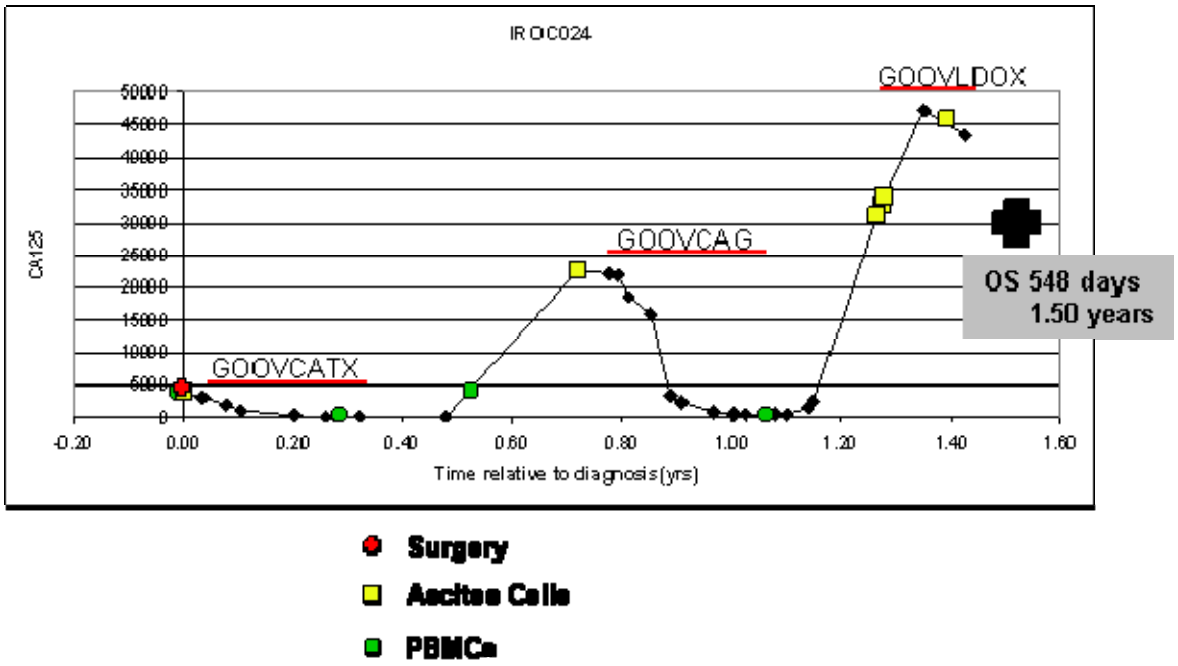



**Figure 1.** Clinical history and available specimens for IROC024, a patient with high-grade serous ovarian cancer. Tumor burden is indicated by CA125 measurements (Y axis) over time (X axis). Also shown are the time points at which ascites cells (yellow squares) and peripheral blood mononuclear cells (PMBC; green squares) were collected. Chemotherapy protocols are shown (GOOVCATX = carboplatin and paclitaxel; GOOVCAG = carboplatin and gemcitibine; GOOVLDOX = pegylated liposomal doxorubicin). Note that ascites cells (containing tumor) were successfully collected at primary surgery, first recurrence, and second recurrence. The patient only survived for 548 days (1.5 years).

As part of our flow cytometry experiments, we have also identified a subpopulation of tumor-infiltrating CD8+ T cells that express the activation marker CD103. We hypothesize that the CD103+ subset is enriched for tumor-reactive T cell clones. For details, please see Figure 1 of Webb et al., 2010 **(Appendix A)**.

In addition, we have developed and published new methods to sequence the TCR repertoire from virtually any human specimen using next-generation sequencing technology (see manuscripts by Warren et al. 2009 and Freeman et al. 2009, **Appendices B and C**). Indeed, we were the first group to apply this sequencing technology to the human T cell repertoire and, in one experiment, we surpassed the number of human TCR sequences identified by all other labs to that point in history.

Finally, we performed some very basic analysis of the effects of chemotherapy on T cells in EOC patients. To our surprise, we found that the absolute lymphocyte count (ALC) changes very little during chemotherapy in most patients. Instead, patients presented with high, medium or low ALC counts, and these levels persisted through treatment. Remarkably, we noted that patients with low ALC values had decreased survival rates **(Figure 2 below)**. We are currently investigating whether high ALC values correlate with tumor-infiltrating CD8+ T cells, which might

offer an explanation for this unexpected finding. These results are being prepared for publication (Milne et al., in preparation).

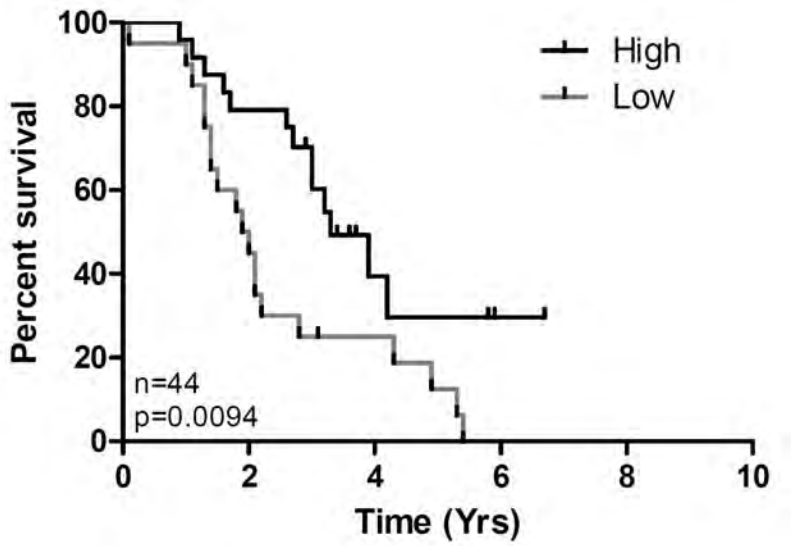



**Figure 2.** Absolute lymphocyte count (ALC) is associated with survival in ovarian cancer. Kaplan-Meier analysis of overall survival for 44 patients with suboptimally debulked high-grade serous ovarian cancer. Patients were stratified into upper (High) and lower (Low) quartiles based on average ALC values recorded during primary chemotherapy. Note that patients with low ALC have significantly decreased survival.

***Task 3. To identify by serological approaches tumor antigens induced by chemotherapy in EOC.***

We proposed to construct autologous cDNA libraries in a yeast-based expression system using mRNA from recurrent ascites tumor cells. Libraries will then be screened with patient sera to identify tumor antigens recognized by IgG autoantibodies. For each antigen, we will determine whether expression is seen in the primary and/or recurrent tumor tissue (by QPCR); whether the antigen shows sequence alterations; and the clinical time point at which antibody responses develop. The overarching goal is to identify new antigens that arise during or after chemotherapy.

***Progress to date:*** Like Task 1, this aim depends on the collection of matched primary and recurrent ascites tumor specimens from patients experiencing a favorable PFI. Therefore, we are still in the anticipated waiting period. In the meantime, however, we have used serological methods to analyze changes in tumor-specific antibodies as patients go through standard treatment. To this end, we assembled a cohort of 23 EOC patients from whom we have serial blood samples spanning from pre-treatment, through to at least one year post-treatment. Initially, we are studying antibody responses using the simple but effective method of Western blotting, as per our prior study in prostate cancer (Nesslinger 2007). Remarkably, 43% of patients (10/23) showed the development of new antibody responses to tumor-associated antigens within 3-12 months of standard treatment **(Figure 3)**.

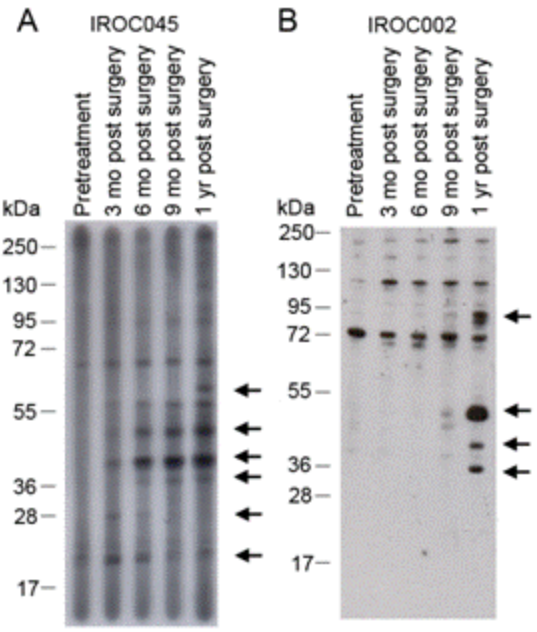



**Figure 3.** Representative immunoblots showing the development of autoantibodies to tumor-associated antigens in ovarian cancer. Serial serum samples were collected at the indicated time points from two patients (A, IROC045 and B, IROC002) with high-grade serous ovarian cancer. Serum was diluted 1:200 and probed against lysate from the human ovarian cancer cell line OVCAR3, followed by anti-human secondary antibody and standard detection by enhanced chemiluminescence. Arrows denote the locations of immunoreactive bands that arise after standard treatment (surgery and chemotherapy). These bands represent antigens against which the patients developed serum autoantibodies.

The above results are consistent with our hypothesis that chemotherapy triggers immune responses against tumors. We are currently attempting to clone the antigens underlying these antibody responses using SEREX methodology, which is similar to that proposed in the grant.

***Task 4. To identify changes to the tumor transcriptome induced by chemotherapy in EOC.***

To identify additional candidate tumor antigens, we proposed to subject primary and recurrent tumor cells to whole-transcriptome cDNA sequencing using a massively parallel sequencing platform (Illumina). Data will then be analyzed to identify sequence alterations or changes in expression level that differ between primary and recurrent tumor tissue using constitutional DNA from matched PBMC as a reference. We expect to identify sequence alterations arising in recurrent tumor cells, including point mutations, deletions and rearrangements. Potential CD8+ T cell epitopes will be identified using computer algorithms predictive of MHC binding.

***Progress to date:*** Like Tasks 1 and 3, this aim depends on the collection of matched primary and recurrent ascites tumor specimens from patients experiencing a favorable PFI. Therefore, we are still in the anticipated waiting period. In the meantime, however, we have

improved our whole-transcriptome cDNA sequencing method. In the original proposal we indicated we would generate approximately 2 billion bp from each of the 12 paired EOC samples. Now, due to improvements in read length and cluster density we have implemented at the Genome Science Centre we have the ability to generate 20 billion bp per library on the Illumina platform, at no additional cost. We will be able to leverage this increased sequencing depth to improve our confidence in mutation identification, and to find mutations in genes with lower expression levels than was initially possible. Further, new methods have recently become available to reliably capture and sequence only coding portions of the genome (ie. the "Exome"), and we are currently validating these methods at our centre. We will explore the possibility of using these two orthogonal approaches, WTSS and Exome sequencing, to increase further the acuity of mutation detection, such that we identify the best candidate epitopes for functional evaluation in the subsequent aim. As described above for Task 1, we are in the process of applying these improved sequencing methods to our first EOC tumor specimen, from IROC024.

***Task 5. To determine whether tumor-associated CD8+ T cells in EOC recognize putative chemotherapy-induced antigens.***

We proposed to use Interferon-γ ELISPOT to test candidate antigens from Tasks 2 and 3 for recognition by CD8+ T cells from recurrent ascites. RNA-transfected or peptide-pulsed CD40L-stimulated B cells will serve as antigen presenting cells. The clinical time point at which antigen-specific T cell responses arise will be assessed by ELISPOT analysis of serial blood samples and primary vs. recurrent ascites samples.

***Progress to date:*** This aim depends on the results of Tasks 2 and 3, therefore it is not expected to commence for 1-2 more years. Nonetheless, we have been developing our ELISPOT platform in preparation for this work. Specifically, we have optimized methods for performing *in vitro* stimulation and expansion of antigen-specific human CD8+ T cells. For this purpose, we have used the melanoma tumor antigen MART-1, which is often used as a model antigen for tumor immunology studies. Our methods are optimized to the point where we can activate and expand naive MART-1-specific CD8+ T cells reliably from normal donors. Thus, we are now prepared to analyze novel tumor antigens as they arise from Tasks 2 and 3.

While ELISPOT is a rapid and sensitive method to quantify T cell responses, it gives only a partial assessment of the phenotype of the T cell responses. A better index may be the so-called "polyfunctional" status of T cells, wherein the cells capacity to produce a spectrum of different cytokines is measured. Therefore, we have optimized the use of multi-parameter flow cytometry to assess T cell polyfunctionality in EOC. Using this method, we have found that T cell polyfunctionality in response to polyclonal stimulation (anti-CD3) is suppressed in the tumor environment of most (but not all) EOC patients, and that this suppression can be reversed by the addition of the stimulatory cytokines IL-2, IL-12 and IL-18 (for example, see **Fig. 3 of Tran et al., Appendix D**).

## KEY RESEARCH ACCOMPLISHMENTS:

- Specimen collection is proceeding as planned, and 7 patients are currently candidates to provide the necessary materials for this study.

- We optimized flow cytometric methods for isolating tumor cells and T cells for massively parallel sequencing. We have also acquired evidence that the CD103+ subset of CD8+ T cells is enriched for tumor-reactive clones (Webb et al., 2010, Appendix A).

- We optimized methods for whole transcriptome shotgun sequencing, and the first patient sample is now being analyzed (Figure 1).

- We found an association between absolute lymphocyte counts (ALC) and survival in EOC (Figure 2) (Milne et al., in preparation).

- We developed an innovative method to analyze the human T cell repertoire using massively parallel sequencing (Warren et al. 2009, Freeman et al. 2009, Appendices B and C).

- We showed that over 40% of EOC patients develop autoantibody responses to tumor-associated antigens during standard treatment (Figure 3).

- We optimized methods for in vitro expansion of human CD8+ T cells for ELISPOT.

- We developed methods for analyzing the functional status of CD8+ T cells by multi-parameter flow cytometry (Tran et al., submitted, Appendix D).

**REPORTABLE OUTCOMES:**

**Published manuscripts:**

1. Nelson BH. IDO and outcomes in ovarian cancer. *Gynecol Oncol.* 2009. Nov;115(2):179-80. PubMed PMID: 19822256.

2. Milne K, Köbel M, Kalloger SE, Barnes RO, Gao D, Gilks CB, Watson PH, Nelson BH. Systematic analysis of immune infiltrates in high-grade serous ovarian cancer reveals CD20, FoxP3 and TIA-1 as positive prognostic factors. *PLoS One*. 2009 Jul 29;4(7):e6412. PubMed PMID: 19641607.

3. Freeman JD, Warren RL, Webb JR, Nelson BH, Holt RA. Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res.* 2009 Oct;19(10):1817-24. Epub 2009 Jun 18. PubMed PMID: 19541912.

4. Warren RL, Nelson BH, Holt RA. Profiling model T-cell metagenomes with short reads. Bioinformatics. 2009 Feb 15;25(4):458-64. Epub 2009 Jan 9. PubMed PMID: 19136549.

5. Warren RL and Holt RA. A census of predicted mutational epitopes for immunological cancer control. Hum. Immunol. 2010 Mar; 71(3):245-54.

6. Webb, J.R., Wick, D.A. Tran, E. Nielsen, J.S., Milne, K., McMurtrie, E. and Nelson B.H. 2010. Expression of the intraepithelial lymphocyte marker $\alpha$E/$\beta$7 Integrin (CD103) is associated with tumor-reactive CD8+ T cells in ovarian cancer malignant ascites. *Gynecol Oncol.* 2010 Sep;118(3):228-36. Epub 2010 Jun 11. PubMed PMID: 20541243.

**Submitted manuscripts:**

1. Tran, E., Nielsen, J.S., Wick, D.A., Ng, A.V., Nesslinger, N.J., McMurtrie, E., Webb, J.R., Nelson, B.H. 2010. Polyfunctional T-cell responses are disrupted by the ovarian cancer ascites environment and only partially restored by clinically relevant cytokines. *Submitted*.

**Leveraged funding:**

Since this grant from the DOD was awarded, we have received funding for the following related projects:

1. Canadian Institutes of Health Research (CIHR) – Grant#:MOP97897
10/2009 – 09/2012
Tumor-infiltrating T cells in ovarian cancer: functional impact on patient survival
Goal: To define the functional phenotype of tumor-infiltrating T cells in ovarian cancer.
PI: Brad Nelson
Co-PI's: John Webb, Peter Watson, Julian Lum

2. Canadian Institutes of Health Research (CIHR) - Grant#: CSB94217
04/2009 – 03/2010
3. The Ovarian Cancer Immune Epitope Database
Goal: To use proteomic methods to identify T cell epitopes for the immunotherapy of ovarian cancer.
PI: John Webb
Co-PI's: Cristoph Borchers, Brad Nelson, Julian Lum

4. Canadian Institutes of Health Research (CIHR) - Grant#: MOP-102679
04/2010 – 03/2014
Characterizing the human T-cell receptor repertoire by massively parallel sequencing
Goal: To characterize individual variation in the human T-cell repertoires at sequence level resolution, using a comparative approach.
PI: Rob Holt
Co-PI: John Webb

**CONCLUSION:**

Overall, this study is progressing on schedule and on budget, with no significant deviations from the original proposal. We are going through the anticipated waiting period associated with the collection of recurrent tumor specimens from patients experiencing a favorable progression-free interval. In the meantime, we have optimized methods for cell sorting of tumor cells, whole transcriptome shotgun sequencing, and in vitro expansion of human T cells for ELISPOT. We have published 5 relevant manuscripts in 2009, and two are under review. Additional, complementary funding has been received or requested from several other agencies, enhancing the strength of our ovarian cancer research program.

**REFERENCES:**
None.

**APPENDICES:**

A. Webb, J.R., Wick, D.A. Tran, E. Nielsen, J.S., Milne, K., McMurtrie, E. and Nelson B.H. 2010. Expression of the intraepithelial lymphocyte marker $\alpha$E/$\beta$7 Integrin (CD103) is associated with tumor-reactive CD8+ T cells in ovarian cancer malignant ascites. *Gynecol Oncol.* 2010 Sep;118(3):228-36. Epub 2010 Jun 11. PubMed PMID: 20541243.

B. Warren RL, Nelson BH, Holt RA. Profiling model T-cell metagenomes with short reads. Bioinformatics. 2009 Feb 15;25(4):458-64. Epub 2009 Jan 9. PubMed PMID: 19136549.

C. Freeman JD, Warren RL, Webb JR, Nelson BH, Holt RA. Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res.* 2009 Oct;19(10):1817-24. Epub 2009 Jun 18. PubMed PMID: 19541912.

D. Tran, E., Nielsen, J.S., Wick, D.A., Ng, A.V., Nesslinger, N.J., McMurtrie, E., Webb, J.R., Nelson, B.H. 2010. Polyfunctional T-cell responses are disrupted by the ovarian cancer ascites environment and only partially restored by clinically relevant cytokines. *Submitted*.

E. Nelson BH. IDO and outcomes in ovarian cancer. *Gynecol Oncol.* 2009. Nov;115(2):179-80. PubMed PMID: 19822256.

F. Milne K, Köbel M, Kalloger SE, Barnes RO, Gao D, Gilks CB, Watson PH, Nelson BH. Systematic analysis of immune infiltrates in high-grade serous ovarian cancer reveals CD20, FoxP3 and TIA-1 as positive prognostic factors. *PLoS One*. 2009 Jul 29;4(7):e6412. PubMed PMID: 19641607.

G. Warren RL and Holt RA. A census of predicted mutational epitopes for immunological cancer control. Hum. Immunol. 2010 Mar; 71(3):245-54.

**SUPPORTING DATA:**
None.

Gynecologic Oncology 118 (2010) 228–236

Contents lists available at ScienceDirect

Gynecologic Oncology

journal homepage: www.elsevier.com/locate/ygyno

# Profound elevation of CD8$^+$ T cells expressing the intraepithelial lymphocyte marker CD103 ($\alpha_E/\beta_7$ Integrin) in high-grade serous ovarian cancer☆

John R. Webb [a,b,*], Darin A. Wick [a], Julie S. Nielsen [a,c], Eric Tran [a,b], Katy Milne [a], Elissa McMurtrie [d], Brad H. Nelson [a,b,c,e]

[a] Trev and Joyce Deeley Research Centre, BC Cancer Agency, Victoria, BC, Canada
[b] Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada
[c] Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada
[d] BC Cancer Agency, Victoria, BC, Canada
[e] Department of Biology, University of Victoria, Victoria, BC, Canada

ARTICLE INFO

*Article history:*
Received 26 March 2010
Available online 11 June 2010

*Keywords:*
T cells
CD8
Integrin
Lymphocytes
Ovarian

ABSTRACT

*Introduction.* Tumor-infiltrating CD8$^+$ T cells are strongly associated with survival in high-grade serous ovarian cancer, but their functional phenotype remains poorly defined. The mucosal integrin CD103 ($\alpha_E/\beta_7$) facilitates the infiltration of T cells into epithelial tissues, including gut and lung mucosa, solid organ allografts, and various epithelial cancers. We reasoned that CD103 might also be expressed by tumor-reactive T cells in ovarian cancer.

*Methods.* Flow cytometry was used to assess the frequency and phenotype of CD103-expressing T cells in primary ascites fluid from 13 patients with high-grade serous ovarian cancer and 2 patients with recurrent disease.

*Results.* We report that a subset of patients with advanced serous ovarian cancer have profoundly elevated frequencies of CD103-expressing CD8$^+$ cells in ascites (between 20% and 70% of CD8$^+$ cells in ascites were CD103$^+$) and that CD103 expression correlated with levels of TGF-β in ascitic fluid. Conversely, CD103 was not expressed on CD4$^+$ cells, even in those patients with very high frequencies of CD8$^+$CD103$^+$ cells. CD8$^+$CD103$^+$ cells were antigen-experienced (CD45RA$^-$CD45RO$^+$CD62L$^{lo}$CCR7$^-$) and of an intermediate (EM2) effector memory phenotype (CD27$^+$CD28$^-$). TCR repertoire analysis indicated significant skewing between CD8$^+$CD103$^-$ and CD8$^+$CD103$^+$ T cell subsets, suggesting the two populations contain distinct antigenic specificities. Lastly, HLA pentamer analysis revealed that one patient in the cohort harbored a high frequency of CD8$^+$ T cells in ascites that were specific for the tumor antigen NY-ESO-1, and that ~75% of these NY-ESO-1 specific CD8$^+$ T cells were CD103$^+$.

*Conclusions.* CD103$^+$ may be a marker of activated and tumor-reactive CD8$^+$ T cells in high-grade serous ovarian cancer.



## Introduction

The integrin CD103 ($\alpha_E/\beta_7$) is expressed on fewer than 2% of circulating peripheral blood cells, but is widely expressed on intraepithelial CD8$^+$ T cells (IEL) present in the gut and lung mucosa [1–3] and on tissue-infiltrating CD8$^+$ T cells during allograft rejection [4]. The only known ligand for CD103 is the epithelial cell surface molecule, E-cadherin, and CD103/E-cadherin interactions are thought to play an important role in the homing and retention properties of intraepithelial lymphocytes [5]. Recently, CD8$^+$CD103$^+$ cells have also been reported to exert regulatory functions via secretion of IL-10 as well as cell contact-mediated suppressive activity [6]. Despite the apparent diversity of activities attributed to CD8$^+$CD103$^+$ cells, all appear to be intimately associated with the availability of TGF-β, which is an integral factor in the regulation of CD103 surface expression [7].

Increasing evidence suggests that interactions between CD103 and E-cadherin also play an important role in specific immunity against cancers of epithelial origin. Specifically, CD103/E-cadherin interactions have been shown to be critically important for recognition and killing of tumor cells by human colon carcinoma-specific CTL [8], lung cancer-specific CTL [9] and pancreatic cancer-specific CTL [10]. In addition, elevated frequencies of CD8$^+$CD103$^+$ tumor-infiltrating cells have been reported in subsets of colorectal cancer [11] and

☆ This work was supported by the British Columbia Cancer Foundation and by a grant to J.R.W. and B.H.N. by the US Department of Defense. J.S.N is supported by a fellowship from the Canadian Institutes of Health Research.

* Corresponding author. Deeley Research Centre, British Columbia Cancer Agency, Victoria, BC, Canada V8R 6V5. Fax: +250 519 2040.
*E-mail address:* jwebb@bccancer.bc.ca (J.R. Webb).

0090-8258/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygyno.2010.05.016

bladder cancer [12]. Perhaps the most intriguing aspect of the association between CD103 and epithelial tumors is the finding that CD103 expression can be induced on otherwise CD103-negative $CD8^+$ T cells by inclusion of TGF-β in culture media during in vitro priming. $CD8^+$ T cells activated in the presence of TGF-β are then primed for rapid CD103 re-expression upon subsequent antigen exposure, even in the presence of very low levels of TGF-β [8]. In addition, CD103 expression by tumor-infiltrating lymphocytes has been shown to play a critical role in retaining tumor-specific lymphocytes within the tumor microenvironment [13]. Thus, CD103 expression by tumor-specific T cells appears to facilitate appropriate homing to relevant tumor sites, which may be important for immunotherapy of epithelial cancers.

Tumor-infiltrating $CD8^+$ T cells are strongly associated with increased progression-free and overall survival in high-grade serous epithelial ovarian cancer (EOC) [14]. Given the above evidence implicating CD103 in the localization and activation of tumor-specific T cells in epithelial cancers, we reasoned CD103 might also be expressed by tumor-associated T cells in ovarian cancer. Herein, we report that a subset of EOC patients have malignant ascites that contains profoundly elevated numbers of $CD8^+CD103^+$ T cells, which are positively correlated with levels of TGF-β in ascites fluid. Phenotypic analysis of these $CD8^+CD103^+$ cells suggests that they are antigen-experienced, effector memory cells with a skewed TCR Vβ repertoire compared to $CD8^+CD103^-$ cells. Furthermore, our data suggests that CD103 may be a marker of tumor antigen-specific $CD8^+$ T cells in EOC, a finding which has implications for the development of effective immune-based treatments for this challenging disease.

## Materials and methods

### *Patients and tissues*

Primary tumor tissue and malignant ascites were obtained from patients with high-grade serous EOC undergoing initial de-bulking surgery. Ascites was also obtained from two patients with recurrent disease. All specimens and clinical data were obtained with written informed consent under protocols approved by the Research Ethics Board of the British Columbia Cancer Agency and the University of British Columbia. Cells in ascites samples were pelleted by centrifugation and analyzed immediately by FACS or were cryopreserved in liquid nitrogen for future analysis. In instances where significant numbers of red blood cells (RBC) were present, samples were treated with ACK lysis buffer (Biowhittaker) prior to centrifugation. After centrifugation, ascites fluid was retained for TGF-β ELISA analysis. Bulk primary tumor tissue was digested overnight at 4 °C with a combination of collagenase, DNase I and hyaluronidase (all from Sigma), then passed through a 100-μM filter, centrifuged and cryopreserved as above.

### *Antibodies and FACS analysis*

Single cell suspensions from malignant ascites or primary tumor tissue were surface stained with antibodies specific for CD4, CD8, CD103, CD25, CD27, CD28, HLA-DR, CCR7 and CD137 as indicated (all from BD Biosciences) and were analyzed on a BD FACSCalibur using a live lymphocyte gate. TCR spectratyping was performed by staining with CD8 and CD103 plus a combination of antibodies specific for 24 different TCR Vβ chains (IOTEST BetaMark, Beckman Coulter). Where indicated, cells were also stained with an HLA-A2 pentamer reagent loaded with the HLA-A2 restricted epitope of NY-ESO-1 (NY-ESO-$1_{157–165}$) (Proimmune) following manufacturer's instructions. For intracellular cytokine analysis, cells were incubated in cRPMI media (RPMI 1640, 10% FBS, 2 mM L-glutamine, 50 μM 2-mercaptoethanol, 10 mM HEPES, and 10 mM sodium pyruvate) for 6 h in the absence of any additional stimulus, or in the presence of PMA (50 ng/ml) plus ionomycin (250 ng/ml). Egress of cytokine from the cell was inhibited by adding 2 μM monensin (GolgiStop, BD Biosciences) for the duration of the incubation. After 6 h incubation, cells were recovered by centrifugation, surface stained with the indicated antibodies, and fixed and permeabilized using Cytofix/Cytoperm (BD Biosciences) according to manufacturer's instructions. Intracellular cytokines were then detected using anti-IFN-γ or biotinylated anti-IL-10 antibodies plus streptavidin-PE (all from BD Biosciences). TGF-β levels in ascites fluid were quantified by ELISA (eBioscience) according to manufacturer's instructions, and results are reported as a combination of active and latent TGF-β.

### *IFN-γ ELISPOT*

ELISPOT plates (MSIP, Millipore) were pre-coated overnight with 10 μg/ml anti-IFN-γ capture antibody (mAb 1-D1K, Mabtech) and then blocked for 2 h at 37 °C with cRPMI. Ascites cells ($3\times10^5$ cells per well) were plated in triplicate in the absence of any stimulus (media only), or in the presence of 10 μg/ml melan$A_{26–35}$ peptide (irrelevant HLA-A2 binding control peptide) or 10 μg/ml NY-$ESO_{157–165}$ peptide (both peptides from Anaspec). After overnight incubation at 37 °C, ELISPOT plates were washed and incubated for 2 h at 37 °C with 1 μg/ml biotinylated anti-IFN-γ antibody (mAb 7-B6-1, Mabtech) followed by development with Vectastain ABC Elite kit and Vectastain AEC substrate reagent according to manufacturer's instructions (Vector Labs).

### *In vitro expansion of ascites T cells*

Single cell suspensions from malignant ascites were incubated for 7 days in cRPMI ($1\times10^6$ cells/ml) in the presence of 6000 U/ml IL-2 (Tecin, Roche) with or without 2 ng/ml TGF-β (Peprotech) as indicated. After 7 days, a portion of the cell culture was used for FACS analysis and the remainder was subjected to rapid expansion protocol (REP). Briefly, REP cultures are comprised of IL-2-expanded cells mixed at a 1:10 ratio with autologous irradiated (3500 rad) PBMC plus anti-CD3 (30 ng/ml, OKT3, eBioscience) and IL-2 (50 U/ml, Tecin, Roche). The REP was performed in the presence or absence of 2 ng/ml TGF-β (Peprotech), as indicated, for 14 days. Cells were then harvested and re-stimulated ($4\times10^5$ cells per ml) with media only, immobilized anti-CD3 (5 μg/ml) or $2\times10^4$ autologous tumor cells for 6 h in the presence of GolgiStop (BD Biosciences) and were analyzed by FACS to detect CD8 and CD103 surface expression and intracellular IFN-γ.

## Results

### *CD103 surface expression on $CD8^+$ T cells in malignant ascites of ovarian cancer patients*

Approximately 30% of patients with high-grade serous EOC present with ascites fluid (malignant ascites), which typically contains a variable mixture of tumor cells and inflammatory cells. To better define the nature of the host immune response to EOC, we collected ascites from 13 previously untreated EOC patients at the time of their primary de-bulking surgery and analyzed resident lymphocyte populations by flow cytometry using a panel of antibodies against lymphocyte surface markers. In some ascites specimens, there was a dramatic accumulation of $CD8^+$ T cells that expressed CD103 ($\alpha_E/\beta_7$ integrin) on the cell surface (Fig. 1A). Similarly, we observed that a high proportion of $CD8^+$ lymphocytes in solid tumor tissue also expressed CD103 on the cell surface (Fig. 1B). The proportion of $CD8^+CD103^+$ cells in ascites varied widely, ranging from approximately 3% (IROC008) to greater than 70% (IROC033) of all $CD8^+$ T cells (Fig. 1C). For three of the patients in the cohort with a high frequency of CD103-expressing $CD8^+$ T cells (IROC024, IROC0033 and IROC065) this corresponded to approximately

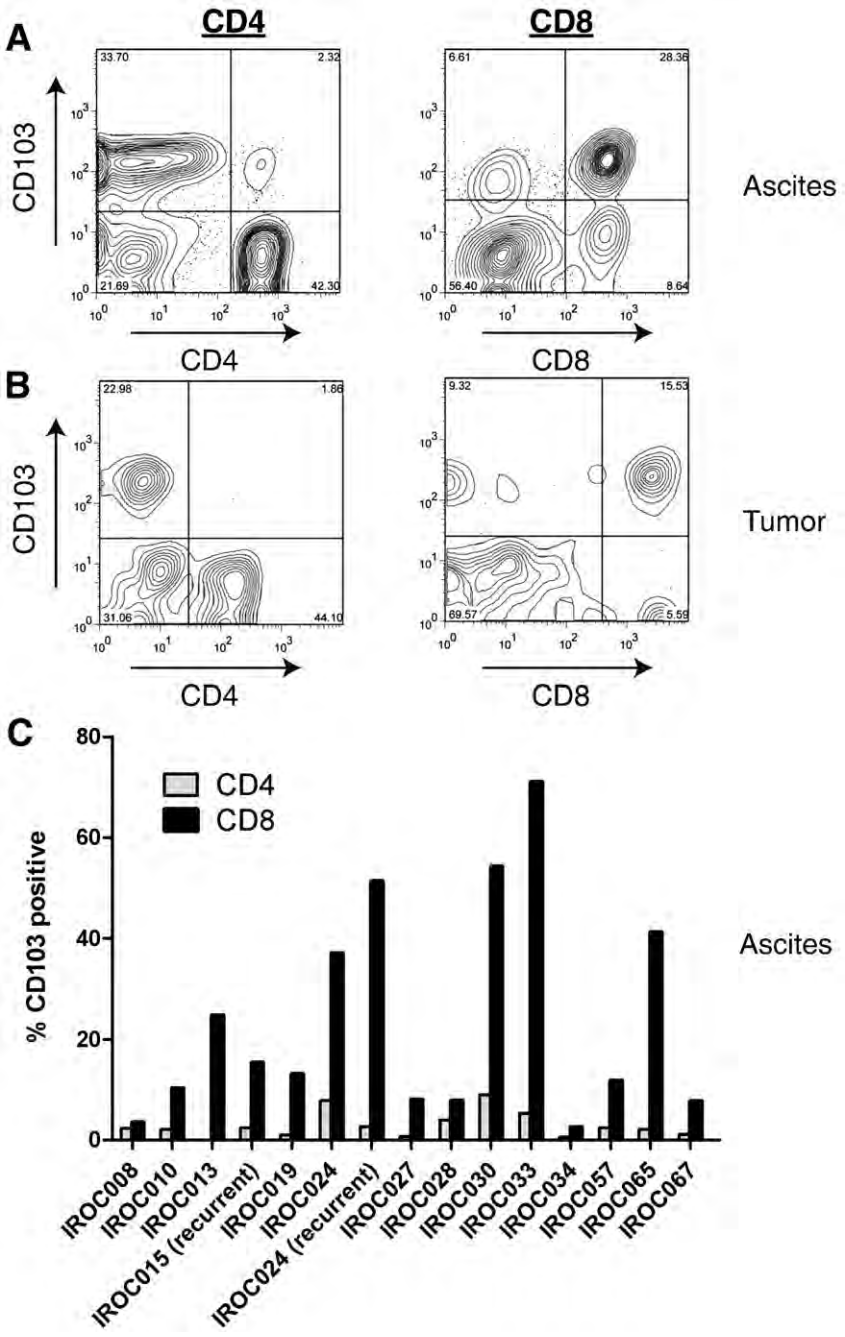



**Fig. 1.** $CD8^+CD103^+$ T cells in malignant ascites and solid tumors of serous EOC patients. (A and B) Flow cytometric analysis showing CD103 expression on $CD4^+$ and $CD8^+$ T cells in (A) ascites or (B) enzymatically dissociated primary tumor from a single patient at the time of initial de-bulking surgery. (C) CD103 expression on $CD4^+$ and $CD8^+$ T cells in ascites samples from 14 serous ovarian cancer patients, including one patient (IROC024) for which primary and recurrent samples (8 months after primary de-bulking) were available. Data in panels A and B are plotted as the percentage of CD103-expressing cells within the total lymphocyte gate whereas data in panel C are plotted as the number of CD103-expressing cells within the CD4 or CD8 lymphocyte population.

$1.6\times10^8$, $3.3\times10^8$ and $1.1\times10^8$ total $CD8^+CD103^+$ cells, respectively, in the volume of ascites fluid recovered during surgery (IROC024-1800 ml, IROC033-900 ml and IROC065-475 ml).

Simultaneous analysis of the CD4 compartment revealed that, unlike $CD8^+$ T cells, only a small minority of $CD4^+$ T cells in malignant ascites, or solid tumor tissue, expressed CD103, even in those patients with very high frequencies of $CD8^+CD103^+$ cells. Interestingly, during the course of the study, one of the patients (IROC024) required a second, palliative paracentesis procedure to remove a subsequent ascites buildup (8 months after primary de-bulking surgery). A large population of $CD8^+CD103^+$ cells was still present in this recurrent ascites sample, suggesting that this lymphocyte subpopulation is stable over time. $CD8^+CD103^+$ cells were also detected in the ascites of a second patient with recurrent disease (IROC015). However this patient did not present with ascites at primary surgery thus no comparison between primary and recurrent disease was possible.

The only known ligand for CD103 is the epithelial cell surface adhesion molecule E-cadherin; therefore, we assessed whether the presence of $CD8^+CD103^+$ T cells correlated with the expression of E-cadherin on tumor cells. Immunohistochemical analysis of paraffin-embedded tumor tissue revealed that E-cadherin was expressed on all tumor specimens (data not shown), despite the varying levels of $CD8^+CD103^+$ T cells. Therefore, the frequency of $CD8^+CD103^+$ T cells in ascites is not correlated to E-cadherin expression by tumors.

Previous studies have indicated that CD103 expression on T cells is upregulated in response to TGF-β but only when TGF-β is delivered in conjunction with a TCR-mediated signal [8,15,16]. We therefore assessed the levels of TGF-β in ascites fluid (Fig. 2). Although TGF-β was not detected in every sample with elevated frequencies of $CD8^+CD103^+$ cells (IROC030 for example), there was a statistically significant correlation between the frequency $CD8^+CD103^+$ T cells and levels of TGF-β in ascitic fluid ($p = 0.005$, Pearson correlation, two-tailed).

### *Phenotype of $CD8^+CD103^+$ T cells in malignant ascites of ovarian cancer patients*

It has been previously demonstrated in other epithelial tumor settings that CD103 plays an important role in recognition and killing of tumor targets [8,9]. Therefore, we assessed whether CD103 expression by malignant ascites $CD8^+$ T cells was associated with an activated (effector) T cell phenotype (Fig. 3). We focused upon three patients with the highest percentage of $CD8^+CD103^+$ cells (IROC024, IROC033 and IROC065). The vast majority of $CD8^+$ T cells were negative for expression of the activation markers CD25 and CD137 and the chemokine receptor CCR7, regardless of whether they belonged to the $CD8^+CD103^-$ or $CD8^+CD103^+$ subset. However, $CD8^+CD103^+$ T cells in all three patients expressed higher levels of HLA-DR on the cell surface than did $CD8^+CD103^-$ cells, consistent with $CD8^+CD103^+$ cells being in an activated state. In addition, $CD8^+CD103^+$ and $CD8^+CD103^-$ T cells could be readily discriminated based upon their pattern of CD27 and CD28 expression. Although all ascites $CD8^+$ T cells were generally positive for CD27 expression, $CD8^+CD103^-$ T cells tended to be $CD28^+$ whereas $CD8^+CD103^+$ T cells were either $CD28^-$ or a mixture of $CD28^+$ and $CD28^-$ cells. This pattern of CD27 and CD28 expression, in combination with other markers of differentiation state ($CD45RO^+$, $CD45RA^-$,$CD62L^-$, data not shown), is consistent with $CD8^+CD103^+$ cells being antigen-experienced and belonging to an intermediate effector memory population previously designated 'EM2' [17].

Prior studies have indicated that CD103 expression can also be a characteristic of $CD8^+$ T cells with regulatory functions [6,18,19], including production of IL-10. Thus, we assessed IL-10 and IFN-γ production by $CD8^+CD103^-$ versus $CD8^+CD103^+$ T cells in a bulk ascites cell preparation containing a mixture of $CD8^+CD103^-$ cells and $CD8^+CD103^+$ T cells (Fig. 4A). In the absence of specific stimulation, $CD8^+CD103^+$ T cells appeared to constitutively express a low amount of IL-10 whereas $CD8^+CD103^-$ T cells produced neither IL-10 nor IFN-γ (Fig. 4B). When activated with a polyclonal stimulus (PMA + ionomycin), both the $CD8^+CD103^-$ and $CD8^+CD103^+$ T cells produced IFN-γ. In addition, $CD8^+CD103^+$ T cells continued to produce IL-10 in response to PMA + ionomycin, however the level of IL-10 did not increase beyond constitutive levels. It is interesting to note that although this experiment was performed with a bulk ascites preparation containing a mixture of $CD8^+CD103^-$ and IL-10-producing $CD8^+CD103^+$ T cells, $CD8^+CD103^-$ cells were still capable of producing IFN-γ when stimulated, suggesting that $CD8^+CD103^+$ T cells did not fully abrogate the effector function of neighboring $CD8^+CD103^-$ cells. However, PMA + ionomycin is a potent stimulus that may override more subtle inhibitory effects.

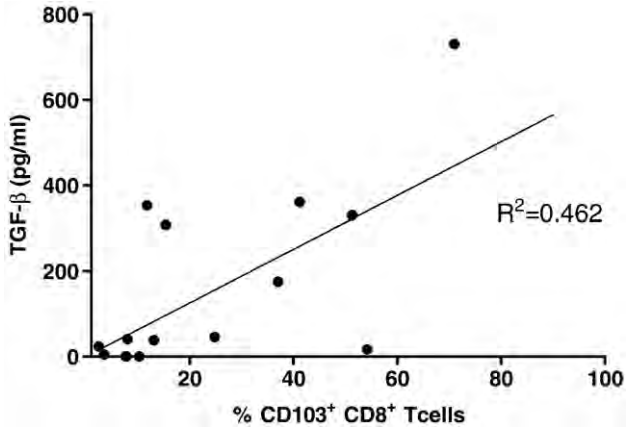



**Fig. 2.** The percentage of $CD8^+CD103^+$ T cells is positively correlated with TGF-β levels in ascites fluid. Ascites collected from patients described in Fig. 1 was centrifuged to remove cellular components and the level of TGF-β in ascites fluid was measured by cytokine capture ELISA. Data are reported as the combination of latent plus active TGF-β (mean of triplicate wells plus SEM) (*Y*-axis) and the percentage of $CD103^+$ $CD8^+$ T cells (*X*-axis) for individual patients from Fig. 1.

### *Comparison of TCR Vβ usage by $CD8^+CD103^+$ and $CD8^+CD103^-$ T cell subsets in malignant ascites*

To further assess whether $CD8^+CD103^+$ cells might constitute tumor-reactive cells, we looked for evidence of skewing of the T cell receptor (TCR) repertoire in this T cell subpopulation. Ascites samples from patients IROC024, IROC033 and IROC065 were stained with CD8- and CD103-specific antibodies in combination with a panel of multiplexed antibodies specific for 24 different human TCR Vβ chains, which together account for approximately 70% of the human TCR Vβ gene repertoire. As shown in Fig. 5, the $CD8^+CD103^-$ and $CD8^+CD103^+$ T cell subpopulations showed distinct patterns of TCR Vβ usage. For example, in patient IROC024, two predominant Vβ subsets (Vβ4 and Vβ7.2) comprised approximately 36% of the $CD8^+CD103^+$ subset, but only 6% of the $CD8^+CD103^-$ subpopulation. Likewise, in patient IROC033, cells expressing Vβ1 and Vβ20 together comprised approximately 26% of the $CD8^+CD103^+$ subset, but less than 6% of the $CD8^+CD103^-$ subpopulation. Even more strikingly, in patient IROC065 the $CD8^+CD103^+$ subset was comprised almost exclusively (>85%) of cells expressing Vβ17. Together these results imply that, based upon their distinct TCR repertoires, the $CD8^+CD103^+$ and $CD8^+CD103^-$ T cell subsets have distinct antigen specificities.

### *$CD8^+$ T cells specific for the tumor antigen NY-ESO-1 fall predominantly in the $CD8^+CD103^+$ subset*

Unlike other tumor sites such as melanoma, there are relatively few defined T cell antigens in EOC, which limits the ability to characterize tumor antigen-specific T cells. However, we and others have reported that a subset of EOC patients has robust $CD8^+$ T cell responses to the cancer-testis antigen NY-ESO-1 [20]. In the present cohort, one $HLA\text{-}A2^+$ patient (IROC013) demonstrated a strong $CD8^+$ T cell response to an HLA-A2-restricted epitope of NY-ESO-1 (NY-ESO-$1_{157-165}$). This response was readily detected by ELISPOT analysis of ascites-derived T cells without the need for ex vivo expansion (Fig. 6A). This same patient also had a high frequency of $CD8^+CD103^+$ T cells in malignant ascites (Fig. 6B). We hypothesized that NY-ESO-1-reactive $CD8^+$ T cells might be restricted to the $CD103^+$ subset. Indeed, by labeling T cells with HLA-A2/NY-ESO-$1_{157-165}$ pentamers, we found that ~75% of NY-ESO-$1_{157-165}$ reactive $CD8^+$ T cells expressed CD103 (Fig. 6C).

### *Co-engagement of TGF-β receptor and T cell receptor is required for induction and maintenance of CD103 expression by ascites-derived $CD8^+$ T cells*

To directly test whether TGF-β could induce CD103 expression on $CD8^+$ T cells and to assess the requirement for TCR co-engagement, T cells from malignant ascites of patients with either a low (IROC008) or high (IROC033) initial frequency of $CD8^+CD103^+$ T cells (Figs. 7A and B, respectively) were expanded in vitro in the presence or absence of TGF-β. Bulk ascites cells were subjected to a conventional ex vivo TIL expansion protocol comprised of an initial round of T cell expansion in high dose IL-2 (6000 U/ml) followed by a subsequent round of rapid expansion (REP) using anti-CD3 antibody (OKT-3) [21]. Both phases of

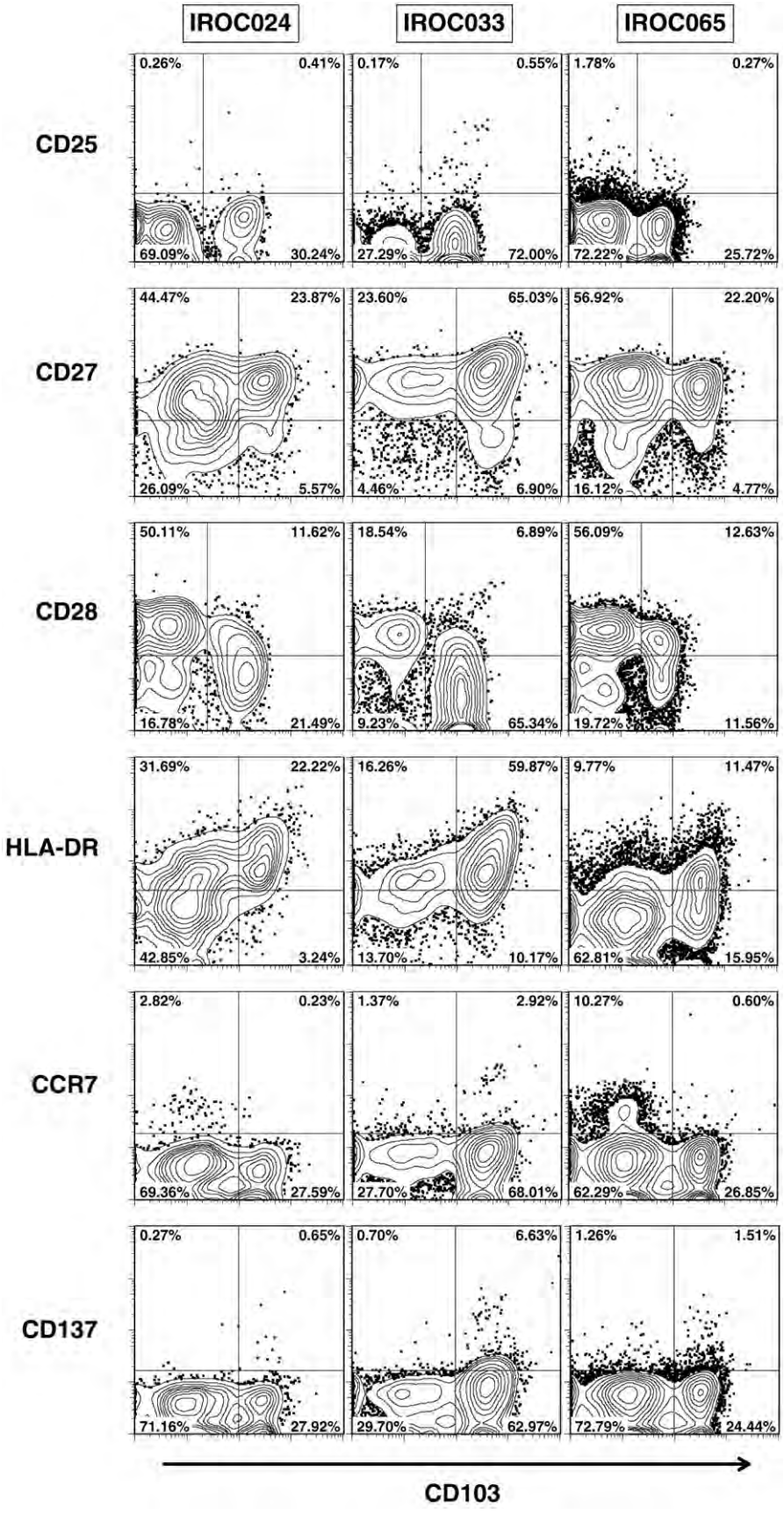



**Fig. 3.** Surface phenotype of $CD8^+CD103^+$ T cells. Malignant ascites samples from three patients with elevated frequencies of $CD8^+CD103^+$ T cells (IROC024, IROC033 and IROC065) were surface labeled with antibodies to CD8 (gated, not shown), CD103 (*X*-axis), and the indicated T cell activation markers (*Y*-axis) and analyzed by flow cytometry. Note that $CD8^+CD103^-$ and $CD8^+CD103^+$ T cells showed distinct expression patterns for CD28 and HLA-DR ($CD8^+CD103^-$ cells were $CD28^+$ and HLA-$DR^{lo}$, whereas $CD8^+CD103^+$ cells were $CD28^{+/-}$ and HLA-$DR^{hi}$.

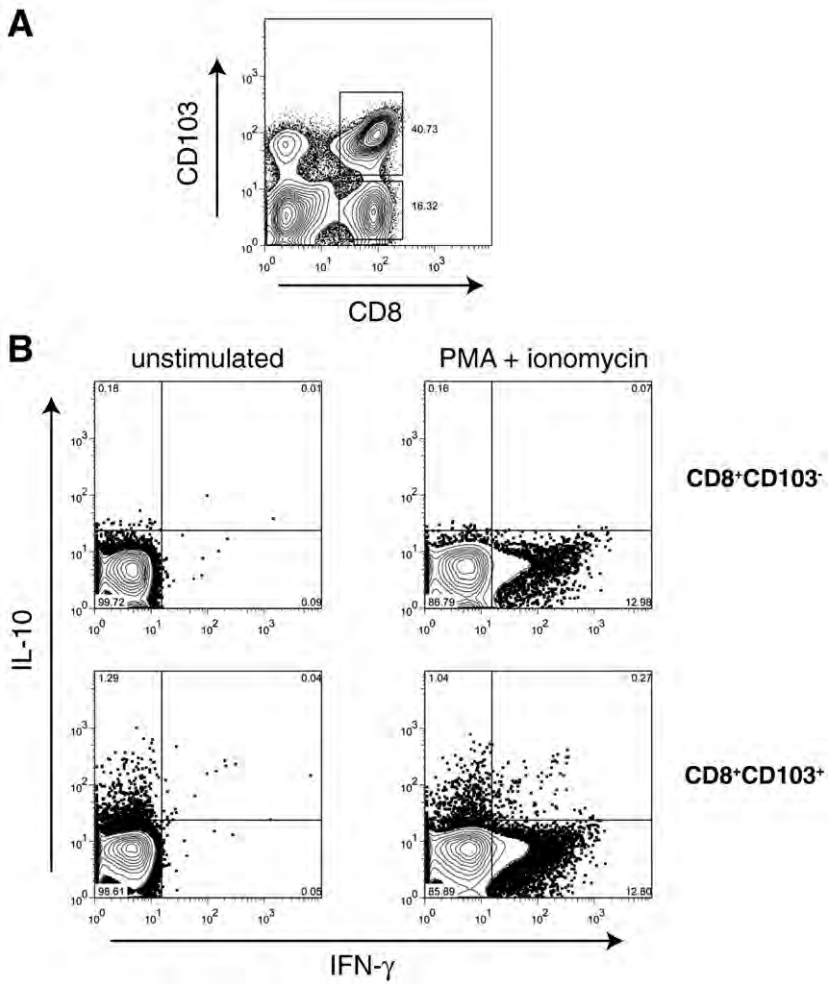



**Fig. 4.** Cytokine production by $CD8^+CD103^+$ T cells. Malignant ascites cells from a representative EOC patient with elevated frequencies of $CD8^+CD103^+$ T cells (IROC033) were analyzed by flow cytometry using antibodies specific for CD8 and CD103 (A), and by intracellular cytokine staining using antibodies specific for IFN-γ and IL-10 (B). In panel B, cells were analyzed after incubation in media only (unstimulated) or after 6 h of stimulation with polyclonal stimulus (PMA + ionomycin). Samples were gated on lymphocyte populations by forward and side scatter. In panel B samples were further gated into $CD8^+CD103^-$ and $CD8^+CD103^+$ subpopulations.

expansion were performed in the presence or absence of exogenous TGF-β. $CD8^+$ T cells retained their initial CD103 phenotype during the IL-2-mediated phase of expansion, regardless of whether or not TGF-β was present in the medium (data not shown). However, $CD8^+$ T cells derived from the ascites sample with a high initial $CD8^+CD103^+$ frequency (IROC033) retained $CD103^+$ expression when expanded with anti-CD3 in the presence of TGF-β (Fig. 7D, right column), but rapidly lost CD103 when expanded with anti-CD3 in the absence of TGF-β (Fig. 7D, left column). In the ascites sample that initially had low proportions of $CD8^+CD103^+$ T cells (IROC008), the addition of exogenous TGF-β during anti-CD3 expansion resulted in the emergence of a population of $CD8^+CD103^+$ T cells (Fig. 7C, right column) whereas cells remained $CD103^-$ when stimulated in the absence of TGF-β (Fig. 7C, left column). Together, these data demonstrate that TGF-β is a key regulator of CD103 expression on ascites $CD8^+$ T cells, but that TGF-β-dependent up-regulation of CD103 surface expression is highly dependent upon concurrent co-stimulation through the TCR.

Lastly, to directly assess whether $CD103^+$ expression may be a potential marker of tumor reactive $CD8^+$ T cells in ascites, we then compared the in vitro expanded T cells from IROC008 versus IROC033 for their ability to recognize autologous tumor cells. After a 14-day period of in vitro expansion, $CD8^+$ T cells from both IROC008 and IROC033 became quiescent and were negative for intracellular IFN-γ staining in the absence of stimulation (Figs. 7C and D, upper panels) and were positive for intracellular IFN-γ staining when re-stimulated with anti-CD3, regardless of whether they were expanded in the presence or absence TGF-β, or their CD103 status (Figs. 7C and D, middle panels). However, when stimulated with autologous tumor cells, only those T cells derived from the ascites sample that initially had high $CD8^+CD103^+$ T cells in vivo (IROC033) produced IFN-γ (Figs. 7C and D, lower panels). Thus, even though CD103 expression could be induced on $CD8^+$ T cells from both patients by ex vivo expansion with anti-CD3 plus TGF-β, only the ascites sample that originally contained $CD8^+CD103^+$ T cells was tumor-reactive.

## Discussion

We report that a subset of patients with high-grade serous EOC has profoundly elevated numbers of $CD8^+CD103^+$ T cells in malignant ascites. $CD8^+CD103^+$ T cells exhibit an antigen-experienced, effector memory phenotype and utilize a distinct TCR Vβ repertoire compared to $CD8^+CD103^-$ T cells. In addition, T cells from a malignant ascites sample with known reactivity to the defined tumor antigen NY-ESO-1 were predominantly of the $CD8^+CD103^+$ subset. To our knowledge, this is the first report that $CD8^+CD103^+$ T cells are present in ovarian cancer and, indeed, in the ascites compartment under any pathological condition. Furthermore, our results suggest CD103 may serve as a marker for isolating tumor-reactive T cells for immunotherapy.

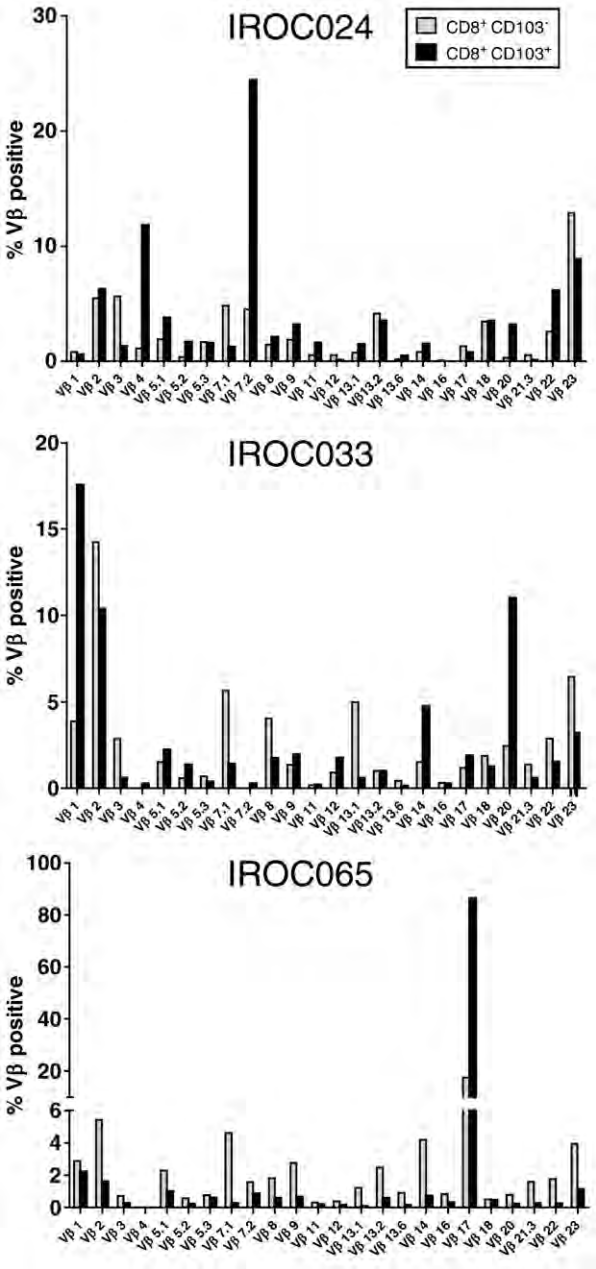



**Fig. 5.** TCR Vβ repertoire of $CD8^+CD103^-$ and $CD8^+CD103^+$ T cell populations. Malignant ascites samples from three EOC patients with elevated frequencies of $CD8^+CD103^+$ T cells (IROC024, IROC033 and IROC065) were surface labeled with antibodies to CD8 and CD103 followed by a panel of multiplexed antibodies specific for 24 different TCR Vβ family members. Data are shown as the percentage of cells expressing a specific TCR Vβ family member after gating into $CD8^+CD103^-$ and $CD8^+CD103^+$ subpopulations.

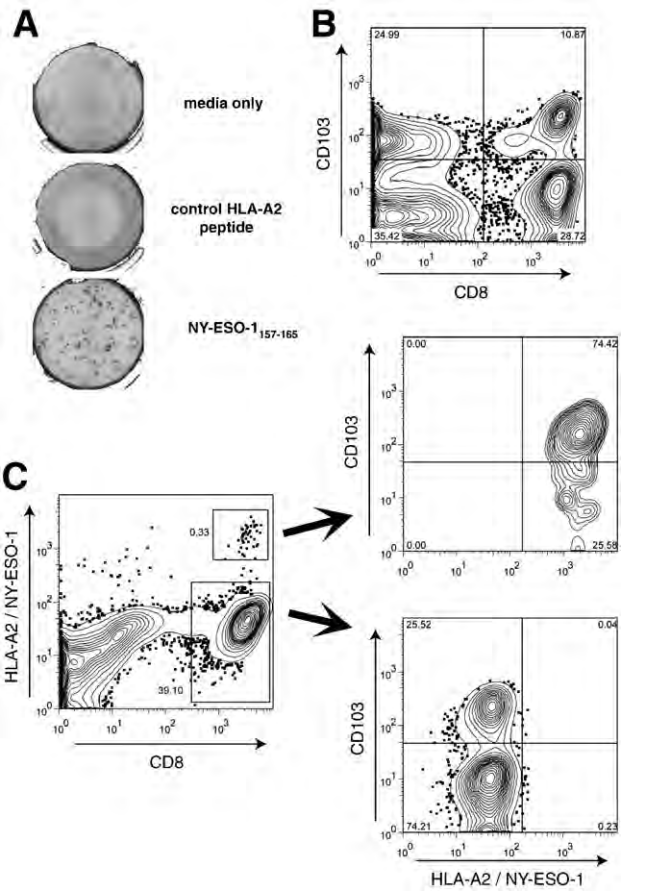



**Fig. 6.** $CD8^+$ T cells specific for the tumor antigen NY-ESO-1 are predominantly $CD103^+$. Panels A–C show results for T cells derived from the malignant ascites of an $HLA\text{-}A2^+$ EOC patient (IROC013) who demonstrated reactivity to an HLA-A2-restricted epitope of NY-ESO-1 (NY-ESO-$1_{157–165}$). (A) Representative single wells (of triplicates) from an IFN-γ ELISPOT assay in which bulk malignant ascites cells were incubated overnight with NY-ESO-$1_{157–165}$ peptide. Negative controls include media alone, or an unrelated HLA-A2-binding peptide (Melan $A_{26–35}$). (B) Flow cytometric analysis showing the frequency of $CD8^+CD103^+$ T cells. (C) Flow cytometric analysis showing the frequency of $CD8^+$ lymphocytes staining positive with an HLA-A2/NY-ESO-$1_{157–165}$ pentamer (Proimmune). Pentamer-positive and -negative $CD8^+$ T cells were further characterized for CD103 expression (right panels). All events in panels B and C were first gated on total lymphocyte populations by forward and side scatter.

Expression of CD103 is normally restricted to $CD8^+$ intraepithelial T cells (IEL) present in mucosal surfaces, where it mediates interaction with epithelial cells via binding to its only known ligand, E-cadherin [1,2]. CD103 mediates interactions between T lymphocytes and epithelial cells via several mechanisms: 1) by facilitating cell to cell adhesion [22,23], 2) through CD103-mediated signaling to the T cell to promote T cell proliferation and effector function [9,24], and 3) by E-cadherin-mediated signaling to the epithelial cell [25]. CD103 is also expressed on mucosal dendritic cells [26,27] and on some $CD4^+CD25^+$ T regulatory cells [28,29].

The prognostic significance of the $CD8^+CD103^+$ T cell population in EOC ascites is currently unknown and will require analysis of a larger patient cohort with longer follow-up. Nonetheless, there is abundant evidence that the presence of intra-tumoral $CD8^+$ T cells in general, and intra-epithelial $CD8^+$ T cells in particular, are associated with long term survival of serous EOC patients [30–32]. Notably, in these studies intraepithelial T cells have been characterized based on their histological location in tumor tissue, rather than on the basis of specific surface markers. Thus, little is known regarding their antigen specificity or tumor homing properties. Unfortunately, there is currently no anti-CD103 antibody available that can be used for paraffin-embedded, formalin-fixed tissue, making it difficult to compare the $CD8^+CD103^+$ T cells identified herein by flow cytometry with intraepithelial $CD8^+$ T cells identified in large retrospective immunohistochemical studies. However, as shown in Fig 1, the relative frequency of CD103-expressing CD8 T cells in bulk ascites and in primary tumor tissue appears to be similar. Furthermore, the results reported here concerning the activation phenotype, TCR Vβ repertoire, and tumor antigen specificity of $CD8^+CD103^+$ T cells in ascites strongly support the hypothesis that CD103 may demarcate the tumor-reactive intraepithelial $CD8^+$ T cell subset associated with long-term survival. If this hypothesis is correct, then these cells warrant further study for antigen discovery and immunotherapy.

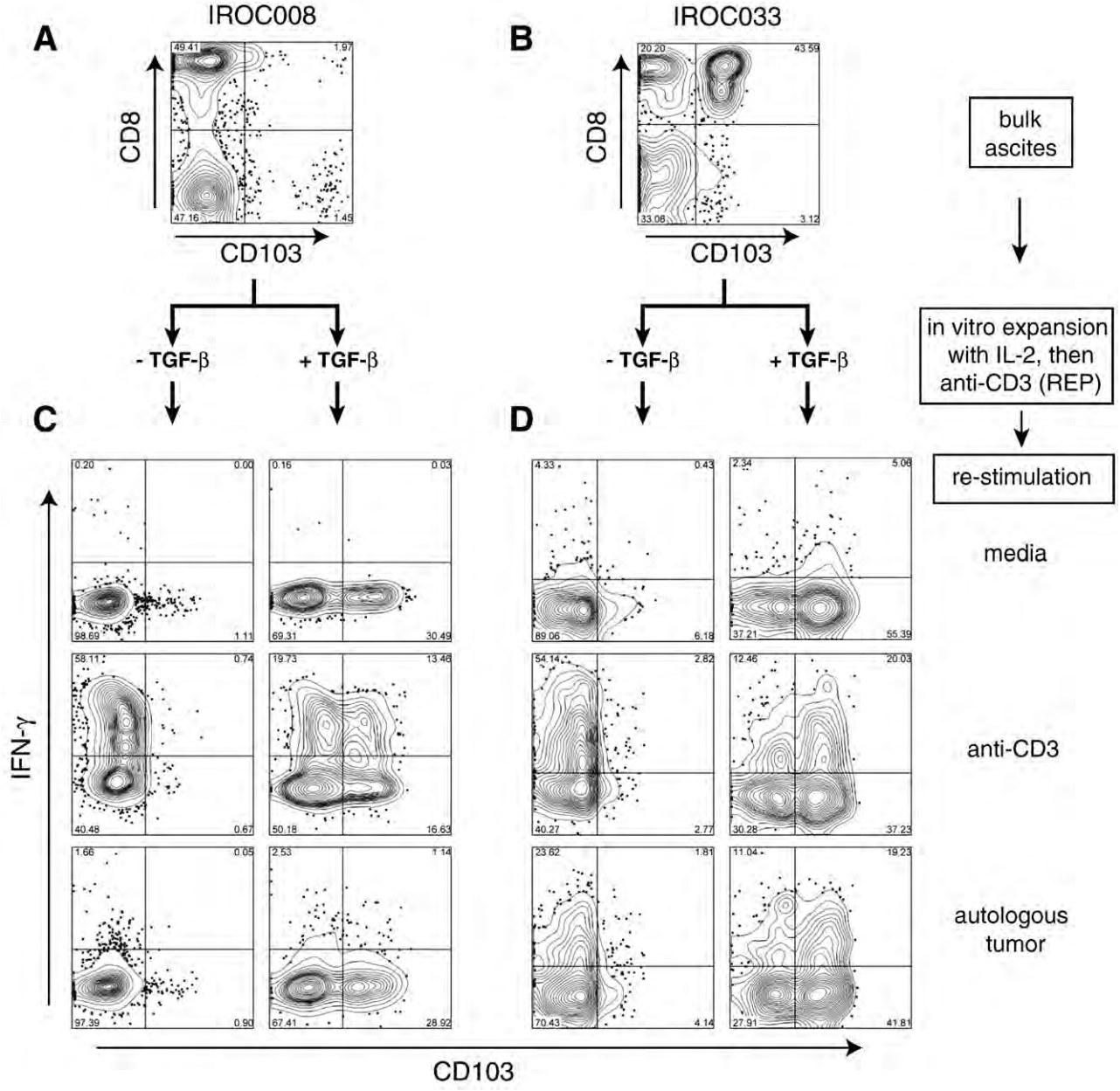



**Fig. 7.** CD103 surface expression on $CD8^+$ T cells can be regulated by inclusion of TGF-β in culture media during the TCR-dependent phase of ex vivo TIL expansion. (A) Bulk ascites cells from two patients with either low (IROC008) or high (IROC033) frequencies of $CD8^+CD103^+$ T cell populations were expanded using a conventional ex vivo TIL expansion protocol comprised of an initial round of expansion in high dose IL-2 followed by a subsequent rapid expansion using anti-CD3 antibody. Both the IL-2 and REP expansions were performed in the presence or absence of exogenous TGF-β (2 ng/ml). At the end of the rapid expansion period, cells were analyzed by intracellular cytokine staining to assess the ability of $CD8^+CD103^-$ and $CD8^+CD103^+$ T cells to produce IFN-γ after re-stimulation with media only, polyclonal stimulation (immobilized anti-CD3 antibody) or autologous tumor cell lines. All events shown in C and D are first gated on $CD8^+$ lymphocyte populations.

Prior studies have shown that TGF-β can upregulate CD103 expression on T cells only if delivered in conjunction with a TCR-dependent signal [8,16] and that such dual signals normally promote IEL retention in mucosal environments [22,33]. These data provide further evidence supporting our hypothesis that $CD8^+CD103^+$ T cells represent tumor-specific lymphocytes, as CD103 expression would only be expected to be elicited by concurrent stimulation with TCR and TGF-β. Indeed we demonstrate herein that by comparing TIL cultures derived from 2 EOC ascites with high and low frequencies of $CD8^+CD103^+$ T cells, only the TIL culture derived from the CD103 high ascites sample was capable of recognizing autologous tumor. A recent study by Ling et al. [8] shows that the presence of TGF-β may be particularly important during the initial priming of naïve T cells, and that T cells primed in the presence of TGF-β require much lower doses of TGF-β to upregulate CD103 during subsequent antigen exposures. These data suggest that TGF-β could potentially be used in TIL cultures to upregulate CD103 expression by $CD8^+$ T cells, thereby enhancing their homing to epithelial tumors after adoptive transfer. Interestingly, despite the fact that TGF-β is widely regarded as an immunosuppressive cytokine, a recent study demonstrated that TGF-β can actually augment the preferential outgrowth of melanoma-reactive cells during the in vitro expansion of TIL for adoptive immunotherapy [34]. Although CD103 was not measured in that study, based upon our results we anticipate that many of the melanoma-reactive CD8 T cells expanded in this protocol may have expressed CD103.

In conclusion, we have demonstrated that the malignant ascites from a significant proportion of EOC patients contains a high frequency of $CD8^+CD103^+$ T cells, which bear the hallmarks of

naturally arising, tumor-specific lymphocytes. These cells warrant further investigation to enhance our understanding of the immunobiology, prognosis and treatment of this devastating disease.

**Conflict of interest statement**
None of the authors have any conflicts of interest.

## Acknowledgments

The authors thank all members of the British Columbia Cancer Agency Tumor Tissue Repository for assistance in patient sample acquisition, Kristy Dillon and Alvin Ng for assistance with patient sample processing and Nancy Nesslinger for critical review of the manuscript.

## References

[1] Cerf-Bensussan N, Jarry A, Brousse N, Lisowska-Grospierre B, Guy-Grand D, Griscelli C. A monoclonal antibody (HML-1) defining a novel membrane molecule present on human intestinal lymphocytes. Eur J Immunol 1987;17:1279–85.
[2] Kruschwitz M, Fritzsche G, Schwarting R, Micklem K, Mason DY, Falini B, et al. Ber-ACT8: new monoclonal antibody to the mucosa lymphocyte antigen. J Clin Pathol 1991;44:636–45.
[3] Lohmeyer J, Friedrich J, Grimminger F, Maus U, Tenter R, Morr H, et al. Expression of mucosa-related integrin alphaEbeta7 on alveolar T cells in interstitial lung diseases. Clin Exp Immunol 1999;116:340–6.
[4] Hadley GA, Bartlett ST, Via CS, Rostapshova EA, Moainie S. The epithelial cell-specific integrin, CD103 (alpha E integrin), defines a novel subset of alloreactive CD8+ CTL. J Immunol 1997;159:3748–56.
[5] Schon MP, Arya A, Murphy EA, Adams CM, Strauch UG, Agace WW, et al. Mucosal T lymphocyte numbers are selectively reduced in integrin alpha E (CD103)-deficient mice. J Immunol 1999;162:6641–9.
[6] Uss E, Rowshani AT, Hooibrink B, Lardy NM, van Lier RA, ten Berge IJ. CD103 is a marker for alloantigen-induced regulatory CD8+ T cells. J Immunol 2006;177:2775–83.
[7] Parker CM, Cepek KL, Russell GJ, Shaw SK, Posnett DN, Schwarting R, et al. A family of beta 7 integrins on human mucosal lymphocytes. Proc Natl Acad Sci U S A 1992;89:1924–8.
[8] Ling KL, Dulphy N, Bahl P, Salio M, Maskell K, Piris J, et al. Modulation of CD103 expression on human colon carcinoma-specific CTL. J Immunol 2007;178:2908–15.
[9] Le Floc'h A, Jalil A, Vergnon I, Le Maux Chansac B, Lazar V, Bismuth G, Chouaib S, Mami-Chouaib F. Alpha E beta 7 integrin interaction with E-cadherin promotes antitumor CTL activity by triggering lytic granule polarization and exocytosis. J Exp Med 2007;204:559–70.
[10] French JJ, Cresswell J, Wong WK, Seymour K, Charnley RM, Kirby JA. T cell adhesion and cytolysis of pancreatic cancer cells: a role for E-cadherin in immunotherapy? Br J Cancer 2002;87:1034–41.
[11] Quinn E, Hawkins N, Yip YL, Suter C, Ward R. CD103+ intraepithelial lymphocytes—a unique population in microsatellite unstable sporadic colorectal cancer. Eur J Cancer 2003;39:469–75.
[12] Cresswell J, Robertson H, Neal DE, Griffiths TR, Kirby JA. Distribution of lymphocytes of the alpha(E)beta(7) phenotype and E-cadherin in normal human urothelium and bladder carcinomas. Clin Exp Immunol 2001;126:397–402.
[13] Franciszkiewicz K, Le Floc'h A, Jalil A, Vigant F, Robert T, Vergnon I, et al. Intratumoral induction of CD103 triggers tumor-specific CTL function and CCR5-dependent T-cell retention. Cancer Res 2009;69:6249–55.
[14] Milne K, Kobel M, Kalloger SE, Barnes RO, Gao D, Gilks CB, et al. Systematic analysis of immune infiltrates in high-grade serous ovarian cancer reveals CD20, FoxP3 and TIA-1 as positive prognostic factors. PLoS ONE 2009;4:e6412.
[15] El-Asady R, Yuan R, Liu K, Wang D, Gress RE, Lucas PJ, et al. TGF-{beta}-dependent CD103 expression by CD8(+) T cells promotes selective destruction of the host intestinal epithelium during graft-versus-host disease. J Exp Med 2005;201:1647–57.
[16] Glader PS, Lofdahl CG, von Wachenfeldt KA. alphaEbeta7 expression on CD8+ T-cells in COPD BAL fluid and on TGF-beta stimulated T-cells in vitro. Lung 2005;183:123–38.
[17] Romero P, Zippelius A, Kurth I, Pittet MJ, Touvrey C, Iancu EM, et al. Four functionally distinct populations of human effector-memory CD8+ T lymphocytes. J Immunol 2007;178:4112–9.
[18] Koch SD, Uss E, van Lier RA, ten Berge IJ. Alloantigen-induced regulatory CD8+CD103+ T cells. Hum Immunol 2008;69:737–44.
[19] Keino H, Masli S, Sasaki S, Streilein JW, Stein-Streilein J. CD8+ T regulatory cells use a novel genetic program that includes CD103 to suppress Th1 immunity in eye-derived tolerance. Invest Ophthalmol Vis Sci 2006;47:1533–42.
[20] Milne K, Barnes RO, Girardin A, Mawer MA, Nesslinger NJ, Ng A, et al. Tumor-infiltrating T cells correlate with NY-ESO-1-specific autoantibodies in ovarian cancer. PLoS ONE 2008;3:e3409.
[21] Dudley ME, Wunderlich JR, Shelton TE, Even J, Rosenberg SA. Generation of tumor-infiltrating lymphocyte cultures for use in adoptive transfer therapy for melanoma patients. J Immunother 2003;26:332–42.
[22] Cepek KL, Parker CM, Madara JL, Brenner MB. Integrin alpha E beta 7 mediates adhesion of T lymphocytes to epithelial cells. J Immunol 1993;150:3459–70.
[23] Cepek KL, Shaw SK, Parker CM, Russell GJ, Morrow JS, Rimm DL, et al. Adhesion between epithelial cells and T lymphocytes mediated by E-cadherin and the alpha E beta 7 integrin. Nature 1994;372:190–3.
[24] Agace WW, Higgins JM, Sadasivan B, Brenner MB, Parker CM. T-lymphocyte–epithelial–cell interactions: integrin alpha(E)(CD103)beta(7), LEEP-CAM and chemokines. Curr Opin Cell Biol 2000;12:563–8.
[25] Watabe M, Nagafuchi A, Tsukita S, Takeichi M. Induction of polarized cell–cell association and retardation of growth by activation of the E-cadherin–catenin adhesion system in a dispersed carcinoma line. J Cell Biol 1994;127:247–56.
[26] Sung SS, Fu SM, Rose CE, Jr., Gaskin F, Ju ST, Beaty SR. A major lung CD103 (alphaE)-beta7 integrin-positive epithelial dendritic cell population expressing Langerin and tight junction proteins. J Immunol 2006;176:2161–72.
[27] Johansson-Lindbom B, Svensson M, Pabst O, Palmqvist C, Marquez G, Forster R, Agace WW. Functional specialization of gut CD103+ dendritic cells in the regulation of tissue-selective T cell homing. J Exp Med 2005;202:1063–73.
[28] Rao PE, Petrone AL, Ponath PD. Differentiation and expansion of T cells with regulatory function from human peripheral lymphocytes by stimulation in the presence of TGF-{beta}. J Immunol 2005;174:1446–55.
[29] Van VQ, Darwiche J, Raymond M, Lesage S, Bouguermouh S, Rubio M, et al. Cutting edge: CD47 controls the in vivo proliferation and homeostasis of peripheral CD4+CD25+ Foxp3+ regulatory T cells that express CD103. J Immunol 2008;181:5204–8.
[30] Zhang L, Conejo-Garcia JR, Katsaros D, Gimotty PA, Massobrio M, Regnani G, et al. Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer. N Engl J Med 2003;348:203–13.
[31] Sato E, Olson SH, Ahn J, Bundy B, Nishikawa H, Qian F, et al. Intraepithelial CD8+ tumor-infiltrating lymphocytes and a high CD8+/regulatory T cell ratio are associated with favorable prognosis in ovarian cancer. Proc Natl Acad Sci U S A 2005;102:18538–43.
[32] Leffers N, Gooden MJ, de Jong RA, Hoogeboom BN, ten Hoor KA, Hollema H, et al. Prognostic significance of tumor-infiltrating T-lymphocytes in primary and metastatic lesions of advanced stage ovarian cancer. Cancer Immunol Immunother 2009;58:449–59.
[33] Agace WW. T-cell recruitment to the intestinal mucosa. Trends Immunol 2008;29:514–22.
[34] Liu S, Etto T, Rodriguez-Cruz T, Li Y, Wu C, Fulbright OJ, et al. TGF-beta1 induces preferential rapid expansion and persistence of tumor antigen-specific CD8+ T cells for adoptive immunotherapy. J Immunother 2010;33:371–81.

BIOINFORMATICS ORIGINAL PAPER Vol. 25 no. 4 2009, pages 458–464
doi:10.1093/bioinformatics/btp010

*Sequence analysis*

# Profiling model T-cell metagenomes with short reads

René L. Warren[1,*], Brad H. Nelson[2] and Robert A. Holt[1]

[1]BC Cancer Agency, Michael Smith Genome Sciences Centre, 675 West 10th Avenue, Vancouver, BC V5Z 1L3 Canada and [2]BC Cancer Agency, Deeley Research Centre, 2410 Lee Ave, Victoria, BC V8R 6V5 Canada

Received on September 26, 2008; revised on November 28, 2008; accepted on January 1, 2009

Advance Access publication January 9, 2009

Associate Editor: Alfonso Valencia

**ABSTRACT**

**Motivation:** T-cell receptor (TCR) diversity in peripheral blood has not yet been fully profiled with sequence level resolution. Each T-cell clonotype expresses a unique receptor, generated by somatic recombination of TCR genes and the enormous potential for T-cell diversity makes repertoire analysis challenging. We developed a sequencing approach and assembly software (immuno-SSAKE or iSSAKE) for profiling T-cell metagenomes using short reads from the massively parallel sequencing platforms.

**Results:** Models of sequence diversity for the TCR β-chain CDR3 region were built using empirical data and used to simulate, at random, distinct TCR clonotypes at 1–20 p.p.m. Using simulated TCRβ (sTCRβ) sequences, we randomly created 20 million 36 nt reads having 1–2% random error, 20 million 42 or 50 nt reads having 1% random error and 20 million 36 nt reads with 1% error modeled on real short read data. Reads aligning to the end of known TCR variable (V) genes and having consecutive unmatched bases in the adjacent CDR3 were used to seed iSSAKE *de novo* assemblies of CDR3. With assembled 36 nt reads, we detect over 51% and 63% of rare (1 p.p.m.) clonotypes using a random or modeled error distribution, respectively. We detect over 99% of more abundant clonotypes (6 p.p.m. or higher) using either error distribution. Longer reads improve sensitivity, with assembled 42 and 50 nt reads identifying 82.0% and 94.7% of rare 1 p.p.m. clonotypes, respectively. Our approach illustrates the feasibility of complete profiling of the TCR repertoire using new massively parallel short read sequencing technology.

**Availability:** ftp://ftp.bcgsc.ca/supplementary/iSSAKE

**Contact:** rwarren@bcgsc.ca

**Supplementary information:** Supplementary methods and data are available at *Bioinformatics* online.

## 1 INTRODUCTION

Recognition of MHC (major histocompatibility complex)-presented antigen by the T-cell receptor (TCR) is a pivotal process in cell-mediated adaptive immunity. A vast TCR repertoire is required to recognize the enormous diversity of potential antigens in the environment. TCRs are heterodimers that consist predominantly (90–99%) of an α and a β subunit (reviewed in Lefranc and Lefranc, 2001), the remainder consisting of γ–δ heterodimers. Each chain (a TCR subunit is typically referred to as a chain) originates from the genetic rearrangement of a variable (V), joining (J) and constant (C) gene segment (Gascoigne *et al*., 1984; Hedrick *et al*., 1984). Rearranged TCRβ DNA also includes a short (12–16 nt) diversity (D) gene segment between the V and J gene (Fig. 1; Kavaler *et al*., 1984). At the molecular level, two main mechanisms contribute to generate the immense TCR sequence repertoire. Akin to immunoglobulins, the combinatorial diversity of TCR arises from the genetic rearrangement of V, D and J gene segments (Sakano *et al*., 1979) and yields $\sim 5.8 \times 10^6$ possible TCRαβ gene combinations (Janeway *et al*., 2001). Further diversity is generated during this rearrangement by an additional mechanism of base addition and deletion at the junction of V, (D) and J segments, and is known as the N-diversity (Huck *et al*., 1988). Addition of nucleotides by terminal deoxynucleotidyl transferases at the V–J (α) or V–D–J (β) junction (Landau *et al*., 1984) occurs at random and is frequently preceded by base deletion at the 3′ end of V, the 5′ end of J and at both ends of D. This junctional diversity alone can generate $\sim 2 \times 10^{11}$ distinct molecules, bringing the number of theoretically possible TCRαβ to $\sim 10^{18}$ (Janeway *et al*., 2001). The actual number of unique T-cell clonotypes in human blood is at least $\sim 10^7$ ($10^6$ β-chains; Arstila *et al*., 1999). The amino acids encoded at the V–(D)–J junction, a region known as the third complementarity determining region (CDR3), are principally responsible for antigen recognition and define unique TCR clonotypes (Gorski *et al*., 1994). Together, these somatic genome alterations create a diverse T-cell metagenome in every individual.

Profiling the cellular immune response to immune challenge by, for example, vaccination, transplantation, infection or cancer provides valuable insights into immune system integrity and function and the efficacy of prophylactic or therapeutic interventions. Unfortunately, the TCR diversity is such that complete characterization of repertoires still represents an enormous challenge. Current profiling methods, developed 15 years ago, analyze TCR β-chain repertoire complexity based on the CDR3 length diversity within Vβ gene families (Gorski *et al*., 1994; Pannetier *et al*., 1993; Penitente *et al*., 2008). Although they provide a global picture of the repertoire, these low-resolution PCR-based spectratypes do not allow specific identification and quantification of individual T-cell clonotypes. DNA sequencing achieves higher resolution, but large-scale sequence profiling has been infeasible previously due to cost. For instance, sampling 1 million clonotypes (i.e. 10-fold coverage of 1 million 150 nt target CDR3 sequences in a single individual) with traditional Sanger sequencing would cost ∼$1.5 M. Newer sequencing technologies

*To whom correspondence should be addressed.

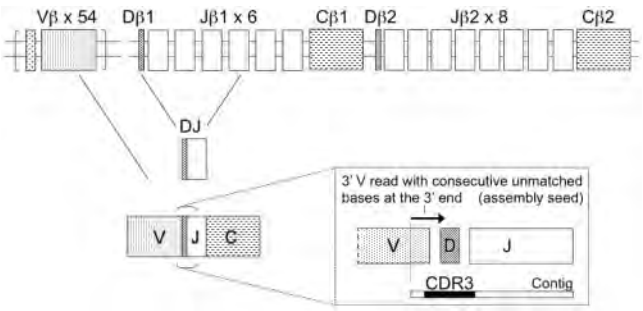



**Fig. 1.** Schematic diagram of the ~1 Mb human TCRβ locus on chromosome 7q34, showing the combinatorial gene rearrangement that takes place and the iSSAKE strategy for assembling CDR3 (inset). The TCRβ locus comprises a cluster of 54 predicted V genes located distantly from two separate clusters each with one D and C gene, interspersed with 6 or 8 J genes. At the DNA level, one of the D genes recombines with one of the J segments, creating partially rearranged DJ genes. Second, one of the V genes joins DJ and the intermediary DNA is deleted. During the gene rearrangements, the random base addition at the junction of V, D and J and the frequent base deletion at the 3′ end of V and 5′ end of the J gene yield the CDR3, a region with unique immune specificities. Read assembly is preceded by the segregation of assembly seeds (arrow); reads that align to the 3′ end of V with eight or more consecutive unmatched 3′ bases. A possible contiguous sequence (contig) resulting from that strategy is shown, with the CDR3-encoding region highlighted in black.

capable of producing a large amount of short reads at much lower cost have emerged in recent years (Bennett, 2004; Holt and Jones, 2008; Margulies *et al*., 2005) and make affordable TCR sequence profiling a likely prospect. Currently, sampling a million TCR clonotypes with the Illumina GAII Analyzer would cost ~1000-fold less compared to Sanger sequencing. On the flip side, next-generation sequencing technologies have much shorter read lengths and show appreciable base error (Holt and Jones, 2008). Combined, these limitations pose a computational challenge for the accurate and complete sequence reconstruction of specificity-determining regions.

Using randomly generated error-prone short reads from simulated TCRβ (sTCRβ) sequences, we have developed a strategy for profiling T-cell metagenomes. The method uses iSSAKE, a modified version of our previously published short read assembler SSAKE (Warren *et al*., 2007), and relies on annotated Vβ gene predictions to segregate partial 3′ alignments and sort corresponding seed sequences prior to assembly (Fig. 1). In this proof-of-principle study, we show that the method is over 63% sensitive for rare 1 p.p.m. clonotypes and over 91% sensitive for clonotypes as low as 2 p.p.m., when using 36 nt reads with 1% randomly distributed errors. When applying a modeled error distribution to simulated reads, we show that the sensitivity of the method is reduced to 51% for the rarest (1 p.p.m.) clonotypes, but is equally sensitive when clonotype frequencies are above 5 p.p.m. The assembly of longer read length impacts positively on the sensitivity of the method. For instance, the method is nearly 12% and 25% more sensitive in recovering 1 p.p.m. sTCRβ, when using 42 nt and 50 nt long reads compared to 36 nt. Together with high base accuracy of over 99%, we show that the majority of CDR3 sequences can be reconstructed accurately and thus characterized using error-rich short read data.

**Table 1.** Frequency (*f*) of base deletion and addition at the CDR3 between publicly available mRNA sequences and simulated TCRβ

| *f* deleted 3′ V bases | | | *f* deleted 5′ J bases | | | *f* added CDR3 bases[a] | | |
|---|---|---|---|---|---|---|---|---|
| Bases | Observed ($N$ = 356) | Simulated ($N$ = 220 000) | Bases | Observed ($N$ = 1151) | Simulated ($N$ = 220 000) | Bases | Observed ($N$ = 174) | Simulated ($N$ = 220 000) |
| 0 | 0.194 | 0.200 | 0 | 0.209 | 0.212 | 1 | 0.006 | 0.007 |
| 1 | 0.160 | 0.158 | 1 | 0.123 | 0.123 | 3 | 0.006 | 0.006 |
| 2 | 0.098 | 0.098 | 2 | 0.122 | 0.122 | 4 | 0.029 | 0.031 |
| 3 | 0.118 | 0.113 | 3 | 0.104 | 0.105 | 5 | 0.017 | 0.019 |
| 4 | 0.160 | 0.155 | 4 | 0.117 | 0.117 | 6 | 0.017 | 0.019 |
| 5 | 0.118 | 0.119 | 5 | 0.123 | 0.119 | 7 | 0.052 | 0.056 |
| 6 | 0.070 | 0.073 | 6 | 0.086 | 0.085 | 8 | 0.063 | 0.067 |
| 7 | 0.045 | 0.047 | 7 | 0.056 | 0.057 | 9 | 0.069 | 0.073 |
| 8 | 0.022 | 0.023 | 8 | 0.031 | 0.031 | 10 | 0.080 | 0.084 |
| 9 | 0.008 | 0.009 | 9 | 0.019 | 0.019 | 11 | 0.057 | 0.059 |
| 10 | 0.006 | 0.006 | 10 | 0.010 | 0.010 | 12 | 0.075 | 0.076 |
| | | | | | | 13 | 0.080 | 0.081 |
| | | | | | | 14 | 0.086 | 0.085 |
| | | | | | | 15 | 0.098 | 0.096 |
| | | | | | | 16 | 0.046 | 0.046 |
| | | | | | | 17 | 0.052 | 0.049 |
| | | | | | | 18 | 0.029 | 0.028 |
| | | | | | | 19 | 0.034 | 0.033 |
| | | | | | | 20 | 0.023 | 0.021 |
| | | | | | | 21 | 0.023 | 0.021 |
| | | | | | | 22 | 0.011 | 0.010 |
| | | | | | | 23 | 0.006 | 0.005 |
| | | | | | | 25 | 0.011 | 0.010 |
| | | | | | | 26 | 0.011 | 0.005 |
| | | | | | | 27 | 0.006 | 0.005 |
| | | | | | | 28 | 0.011 | 0.010 |

[a]We did not observe addition of 2 or 24 nt within the CDR3.

## 2 METHODS

### 2.1 Modeling TCRβ rearrangements

From the alignments of predicted V and J genes to Genbank TCRβ mRNA, four independent models of the N-diversity mechanisms were constructed, representing the frequencies of (i) random base addition at the V–D–J junction; (ii) base composition at each position where bases were added; (iii) 3′ V base deletion; and (iv) 5′ J base deletion. The models are intended as a guide for simulating distinct sequence clonotypes within the CDR3 region by estimating, using empirical data, expected frequencies of base addition, deletion and composition (Table 1). These models were used to construct sTCRβ sequences as described in Supplementary Material. Briefly, this involved, (i) randomly selecting the V and J gene sequences; (ii) deleting 3′ V bases; (iii) deleting 5′ J bases; and (iv) joining V–D–J with addition of junction bases.

For the simulations, we used only the ~150 nt of simulated sequence matching ~40 nt upstream of the 3′ end of V, spanning CDR3 and J and ending ~50 nt downstream of the 5′ end of C. The process of building sTCRβ was repeated to generate a library of 1 000 000 total sequences containing 220 000 unique sTCRβ sequences at frequencies ranging from 1 to 20 p.p.m. (Table 2).

### 2.2 TCRβ–CDR3 reconstruction strategy

From the above 1M sTCRβ sequences, we randomly generated, in three independent replicate experiments, 20 million 36 nt reads having 1.0, 1.5 or 2.0% randomly distributed errors and aligned them to known TCRβ gene

**Table 2.** Generating sTCRβ clonotypes

| Clonotype frequency | p.p.m. | Number of unique sTCRβ | Number of total sTCRβ | Fold coverage ca.[a] |
|---|---|---|---|---|
| 1:1 000 000 | 1 | 110 000 | 110 000 | 5 |
| 1:500 000 | 2 | 10 000 | 20 000 | 10 |
| 1:333 333 | 3 | 10 000 | 30 000 | 15 |
| 1:250 000 | 4 | 10 000 | 40 000 | 20 |
| 1:200 000 | 5 | 10 000 | 50 000 | 25 |
| 1:166 667 | 6 | 10 000 | 60 000 | 30 |
| 1:142 857 | 7 | 10 000 | 70 000 | 35 |
| 1:125 000 | 8 | 10 000 | 80 000 | 40 |
| 1:111 111 | 9 | 10 000 | 90 000 | 45 |
| 1:100 000 | 10 | 10 000 | 100 000 | 50 |
| 1:66 667 | 15 | 10 000 | 150 000 | 75 |
| 1:50 000 | 20 | 10 000 | 200 000 | 100 |

[a]Approximate coverage calculated using 150 nt sTCRβ template size, 20M 36 nt reads.

segments. In addition, we also generated 20M 42 or 50 nt 1% error reads to assess the effect of read length on assembly and tested the effect of random (Dohm *et al*., 2007) versus modeled (Using MAQ simutrain and simulate on real phiX174 Illumina sequences; Heng *et al*., 2008) 1% read error distributions. Simulated reads were aligned against Ensembl (Flicek *et al*., 2008) TCRβ gene predictions using exonerate (Slater and Birney, 2005; Software parameters used: –bestn 1 –score 1 –percent 0). Reads aligning best to V genes, at the 3′ end and having eight or more consecutive unmatched bases in 3′ were put aside as seeds for a *de novo* iSSAKE assembly. The reverse, complemented sequences of reads aligning on the reverse strand of V predictions were also considered as seeds. The assembly read pool consisted of unaligned reads, those aligning to J segments best and assembly seeds. Reads aligning best to V, C or any possible JC junction sequence combinations were discarded (Fig. 1).

### 2.3 Targeted seeded assemblies with iSSAKE

To create the iSSAKE assembler, modifications were made to the SSAKE v3.2.1 code base (Warren *et al*., 2007; http://www.bcgsc.ca/platform/bioinfo/software/ssake). Notably, the depth of the prefix tree was increased, augmenting the number of nodes to 15. This modification was essential to help speed the assembly process at the cost of increased memory requirements. It does so by reducing the search space when considering reads for extension. This modification was compulsory since there is great sequence conservation between the various sTCRβ CDR3 sequences, sometimes differing by only one or a few bases.

Since SSAKE release v2.0 (October 2007), we have implemented the approach for handling error-rich sequencing data described in Jeck *et al*. (2007). In essence, all overhanging bases of reads aligning perfectly to a seed sequence are considered for extension, using a majority rule approach for building consensus sequences of the overhanging bases.

To support the assembly of longer contigs with complete CDR3 without depleting the read pool, sequences used for extension are re-used. The assembly terminates only when all seeds have been maximally extended. This is easily parallelizable on a cluster of computers and permits the assembly of discrete nontruncated TCRβ CDR3 sequences ending in the same J segment. Finally, only the 3′ extension of seeds was permitted, the assembly progressing through V, D and J in this order. For each read set, we ran 100 parallel iSSAKE jobs on two dozen 2.66 GHz Quad-Core 64 bit Intel® Xeon® processors with 15 GB RAM (iSSAKE -m 15 -o 1 -r 0.6).

## 3 RESULTS

A library of 1 million sTCRβ was generated *in silico* using models of TCRβ diversity derived from publicly available mRNA sequences.

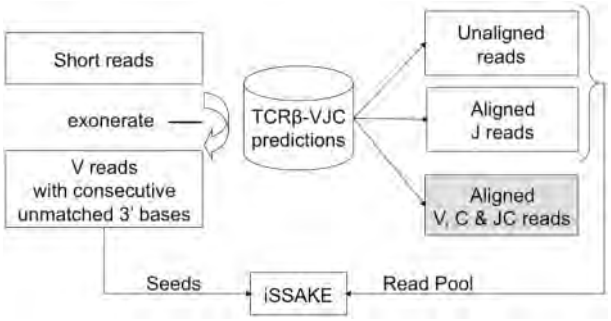



**Fig. 2.** TCRβ–CDR3 reconstruction strategy. Reads are aligned against Ensembl TCRβ gene predictions using exonerate or other short read aligners. Reads aligning best to V genes at the 3′ end and having user defined *n* consecutive unmatched bases in 3′ are set aside as seeds for a *de novo* iSSAKE assembly. The reverse complements of reads aligning on the reverse strand are also considered as seeds. The read pool consists of unaligned reads, those aligning J genes best and assembly seeds. Reads aligning best to V, C or any possible JC junction sequence combinations are discarded to reduce sequence space.

The library consisted of 220 000 distinct sequences present at frequencies ranging from 1 to 20 p.p.m. (110 000 unique 1 p.p.m. sequences and 10 000 unique 2 to 20 p.p.m., Table 2). To confirm that sTCRβ reflect real sequences, we kept track of the simulated N-diversity changes applied to each sequence and verified that the frequencies of V and J base deletion as well as CDR3 base addition in the sTCRβ library were consistent with the frequencies derived from experimental mRNA sequences (Table 1). Twenty million reads having 1.0, 1.5 or 2.0% random or 1.0% modeled error and 36, 42 or 50 nt in length were randomly sampled from the sTCRβ library and assembled in separate experiments. The approximate read coverage for each 1–20 p.p.m. clonotype ranged from 5- to 96-fold, respectively (Table 2).

Since we expect the redundancy of short reads derived from V, J and C to be extremely high and the redundancy over CDR3 to be very low, a classic *de novo* assembly where all sequence reads are used in turn to seed a contig assembly is not suitable. However, the sequences of human TCRβ genes (V, J and C) are known and well annotated, which makes feasible a streamlined strategy of seeded assembly. The assembly seeds we use are sequences that align to the 3′ end of V with eight or more consecutive unmatched 3′ bases in the highly diverse CDR3 region (Fig. 1 inset and Fig. 2). Using exonerate, averages of 1.755, 1.718, 1.679 and 1.599 million seeds were identified from sets of reads with random 1.0, 1.5, 2.0 and 1.0% modeled error, respectively (Table 3). The decrease in number of seeds identified at higher error rates or between random and modeled error distribution (4.3% and 8.9%, respectively) is due to an increased number of mismatched bases that prevent read alignment to the 3′ end of the V gene. Selecting seeds before assembly reduces the sequence space by ∼90% and segregates about half of the 20M input read set for contig assembly. This approach considerably increases the assembly speed and yields only contigs that represent the CDR3.

At 1% randomly distributed error, 84.2 ± 0.16% of the seeds, on average, yielded contigs that comprised complete CDR3 sequences, including D segment bases and unambiguous J segment junctions. These unambiguous contigs which are defined as having clearly

**Table 3.** sTCRβ contig stats from assemblies of 20M randomly generated 36, 42 and 50 nt reads

| Bases | Error (%)[a] | Mean seeds | Number of iSSAKE contigs from triplicates (36 nt) or duplicates (42 and 50 nt) experiments | | | | |
|---|---|---|---|---|---|---|---|
| | | | Short ambiguous[b] | Long ambiguous[c] | Unambiguous[d] | Unambiguous, but sub-optimal[e] | |
| | | | | | | A | B |
| 36 | 1.0[f] | 1 599 140 ± 961 | 498 583 ± 822 | 36 352 ± 261 | 1 064 399 ± 400 | 1098 ± 185 | 2847 ± 89 |
| 36 | 1.0 | 1 755 437 ± 404 | 265 618 ± 245 | 11 520 ± 500 | 1 478 299 ± 236 | 157 ± 19 | 569 ± 5 |
| 36 | 1.5 | 1 718 470 ± 890 | 359 425 ± 882 | 16 460 ± 325 | 1 342 585 ± 1558 | 228 ± 44 | 827 ± 22 |
| 36 | 2.0 | 1 678 905 ± 1018 | 441 317 ± 712 | 22 018 ± 307 | 1 215 570 ± 385 | 302 ± 9 | 1158 ± 53 |
| 42 | 1.0 | 2 448 351 ± 1298 | 369 847 ± 998 | 21 956 ± 315 | 2 056 549 ± 614 | 239 ± 14 | 818 ± 3 |
| 50 | 1.0 | 3 119 789 ± 1150 | 350 076 ± 578 | 341 628 ± 653 | 2 428 086 ± 2380 | 174 ± 12 | 459 ± 8 |

A, misassembled contigs; B, contigs having five or more mismatched bases.
[a] Random error distribution generated using simulators from Dohm *et al*. (2007), unless otherwise specified.
[b] Too short to unambiguously decipher J and thus, CDR3.
[c] Contigs ≥ 45 nt, sufficiently long to contain the first 15 bases of J, but base errors/polymorphisms prevent proper identification of the J segment.
[d] Captured CDR3 and first 15 bases of J unambiguously.
[e] Misassembled contigs are defined here as contigs comprised of reads that belong to distinct sTCRβ. They are identified by looking at discontinuity in the sequence alignment between the contigs and sTCRβ. Contigs having five or more mismatches bases with the closest sTCRβ are built with erroneous reads that often yield misassembled contigs.
[f] Error distribution modeled using phiX174 Illumina sequence data as the training set (Heng *et al*., 2008).

demarcated V and J boundaries were subsequently trimmed to keep the last 15 V bases, the CDR3 and the first 15 bases of identifiable J sequence. The reason for trimming was to facilitate assessment by removing bases that were uninformative for characterization of CDR3. As expected, seeds from 36 nt read sets having higher error rates or errors modeled on Illumina data yield fewer unambiguous contigs, with average proportions of 78.1 ± 0.25% and 72.4 ± 0.10% for the 1.5% and 2% random error sets and 66.6 ± 0.03% for the modeled error read set, respectively (Table 3). This is because random base errors in the seed sequences cause premature termination of contig extension by iSSAKE, unless one or more reads in the pool has matching erroneous bases by chance. The latter case can lead to (i) misassembled contigs, especially if different sTCRβ have a very similar CDR3 makeup and (ii) long ambiguous contigs where J segments are undecipherable.

Misassemblies are identified as contigs that do not match sTCRβ in the source library. These were rarely observed (0.01–0.02% of unambiguous contigs), although more prevalent in the 2% error read set (Table 3). The effect of error on contig misassemblies is more pronounced when error is modeled from real Illumina data (Table 3), where error rates tend to increase toward the end of the read. However, even with a modeled error distribution, misassemblies still represent a minor proportion of all unambiguous reconstructions (0.3%).

We define long ambiguous contigs as those large enough to contain J segment bases, but because of base errors, there was not a precisely matching J segment. An increase from 1% to 2% in the error rate nearly doubles the number of long ambiguous contigs, increasing their proportion from 0.7% to 1.3%. These contigs, while ambiguous, were still useful for assessing the sensitivity of our method. Their abundance is not negligible and the contigs still produce valid alignments to a reference. With real data, identifying the CDR3 from these contigs without a reference sequence will prove more challenging. Short ambiguous contigs are defined here as those that are not long enough to span CDR3. Short ambiguous contigs are caused by early termination of contig extension due to base errors. These short ambiguous contigs represent a considerable portion of the total contigs (15% ± 0.01, 21% ± 0.04, 26% ± 0.05 and 31% ± 0.05 of assemblies using 1.0, 1.5, 2.0% random and 1% modeled error read sets, respectively).

For each assembly, the average base accuracy was calculated by counting the total number of matching bases over the aligned contig length. Although base accuracy of assembled contigs is lower when simulated sequence error rates are higher, it is above 99% at all clonotype frequencies and error rates simulated (Tables 4–6). Contigs representing clonotypes with the lowest frequencies were the least accurate. This is not unexpected since at lower read depths, there are fewer reads to offset the base error, especially in the highly diverse and thus relatively thinly covered CDR3 region. Effectively, inspection of the base error and coverage of assemblies as a function of base position over the region of interest reveals that base mismatch frequency peaks within the seed portion (any of the last 15 V bases and at least 8 consecutive mismatched bases downstream) and decreases through J as the base coverage increases (Fig. 3).

At clonotype frequencies as low as 3 p.p.m., over 93% of the sTCRβ CDR3 sequences could be characterized by iSSAKE contigs assembled from the 1% modeled error distribution read set (Table 4). This means that the sTCRβ sequence diversity can be almost entirely characterized at 15× coverage (Table 2). Although the scope of real T-cell diversity remains unknown, if it is close to the estimated lower limit of $10^6$ β-chains, then substantial repertoire coverage should be easy to attain by massively parallel short read sequencing, even without trimming the reads.

For all contigs that capture sTCRβ sequences we find the accuracy to be very high, especially for clonotypes present at 5 p.p.m. or more. Interestingly, we find that read error has only a small impact on accuracy at these clonotype frequencies. Seeds with errors will rarely find a sequence match in iSSAKE, causing premature contig extension or leading to an increased number of singlets, depleting the pool of unambiguous contigs. Thus, early rejection of these reads has a much more significant impact on the sensitivity than it does on the accuracy.

**Table 4.** Method sensitivity and accuracy as a function of a 1% read error distribution (random or modeled)

| p.p.m. | Number of sTCRβ–CDR3 characterized by iSSAKE contigs (sensitivity) | | Accuracy (%) | | Average number of contigs characterizing each sTCRβ | |
|---|---|---|---|---|---|---|
| | Error (%) | | Error (%) | | Error (%) | |
| | 1.0 Random | 1.0 Modeled | 1.0 Random | 1.0 Modeled | 1.0 Random | 1.0 Modeled |
| 1 | 70 240 | 56 564 | 99.68 | 99.01 | 2.0 | 1.8 |
| 2 | 9111 | 8100 | 99.90 | 99.34 | 3.7 | 2.5 |
| 3 | 9747 | 9295 | 99.96 | 99.64 | 4.6 | 3.4 |
| 4 | 9883 | 9721 | 99.98 | 99.80 | 6.0 | 4.4 |
| 5 | 9932 | 9874 | 99.99 | 99.90 | 7.6 | 5.5 |
| 6 | 9936 | 9913 | 99.99 | 99.94 | 9.1 | 6.6 |
| 7 | 9935 | 9936 | 99.99 | 99.96 | 10.6 | 7.7 |
| 8 | 9939 | 9948 | 99.99 | 99.97 | 12.2 | 8.9 |
| 9 | 9948 | 9955 | 99.98 | 99.97 | 13.7 | 10.0 |
| 10 | 9956 | 9958 | 99.99 | 99.98 | 15.2 | 11.2 |
| 15 | 9972 | 9975 | 99.99 | 99.98 | 23.0 | 16.8 |
| 20 | 9955 | 9958 | 99.98 | 99.98 | 30.7 | 22.1 |

Unambiguous and long ambiguous contigs were used for this analysis. Reported values are the mean of triplicate simulations. Variation among simulations was minimal (Supplementary Table 1).

**Table 5.** Method sensitivity and accuracy as a function of randomly distributed error rates

| p.p.m. | Number of sTCRβ–CDR3 characterized by iSSAKE contigs (sensitivity) | | Accuracy (%) | | Average number of contigs characterizing each sTCRβ | |
|---|---|---|---|---|---|---|
| | Error (%) | | Error (%) | | Error (%) | |
| | 1.5 | 2.0 | 1.5 | 2.0 | 1.5 | 2.0 |
| 1 | 64 862 | 59 259 | 99.48 | 99.23 | 1.9 | 1.8 |
| 2 | 8779 | 8423 | 99.80 | 99.68 | 2.9 | 2.6 |
| 3 | 9656 | 9520 | 99.92 | 99.85 | 4.1 | 3.7 |
| 4 | 9869 | 9831 | 99.96 | 99.93 | 5.4 | 4.9 |
| 5 | 9929 | 9920 | 99.98 | 99.97 | 6.9 | 6.2 |
| 6 | 9936 | 9928 | 99.98 | 99.98 | 8.2 | 7.5 |
| 7 | 9934 | 9924 | 99.98 | 99.97 | 9.7 | 8.7 |
| 8 | 9940 | 9929 | 99.98 | 99.98 | 11.0 | 10.1 |
| 9 | 9952 | 9944 | 99.98 | 99.98 | 12.5 | 11.3 |
| 10 | 9953 | 9940 | 99.99 | 99.98 | 13.9 | 12.6 |
| 15 | 9971 | 9966 | 99.99 | 99.98 | 21.1 | 19.3 |
| 20 | 9961 | 9944 | 99.98 | 99.98 | 28.2 | 26.0 |

Unambiguous and long ambiguous contigs were used for this analysis. Reported values are the mean of triplicate simulations. Variation among simulations was minimal (Supplementary Table 1).

**Table 6.** Method sensitivity and accuracy as a function of read length

| p.p.m. | Number of sTCRβ–CDR3 characterized by iSSAKE contigs (sensitivity) | | Accuracy (%) | | Average number of contigs characterizing each sTCRβ | |
|---|---|---|---|---|---|---|
| | Read length (nt) | | Read length (nt) | | Read length (nt) | |
| | 42 | 50 | 42 | 50 | 42 | 50 |
| 1 | 90 210 | 104 155 | 99.74 | 99.75 | 2.4 | 2.9 |
| 2 | 9737 | 9914 | 99.93 | 99.95 | 4.3 | 5.4 |
| 3 | 9911 | 9959 | 99.98 | 99.98 | 6.4 | 8.0 |
| 4 | 9937 | 9963 | 99.99 | 99.98 | 8.5 | 10.7 |
| 5 | 9948 | 9966 | 99.99 | 99.98 | 10.7 | 13.2 |
| 6 | 9944 | 9966 | 99.98 | 99.98 | 12.8 | 15.8 |
| 7 | 9941 | 9974 | 99.98 | 99.97 | 14.8 | 18.3 |
| 8 | 9948 | 9975 | 99.98 | 99.97 | 17.0 | 20.8 |
| 9 | 9954 | 9979 | 99.98 | 99.98 | 19.1 | 23.2 |
| 10 | 9960 | 9983 | 99.98 | 99.98 | 21.1 | 25.7 |
| 15 | 9973 | 9985 | 99.99 | 99.98 | 31.1 | 37.2 |
| 20 | 9958 | 9976 | 99.98 | 99.98 | 40.0 | 47.4 |

Unambiguous and long ambiguous contigs were used for this analysis. Reported values are the mean of duplicate (42 and 50 nt reads) simulations at 1.0% randomly distributed errors. Variation among simulations was minimal (Supplementary Table 1).

Again, it is important that in real sequence data, errors tend to accumulate toward the 3′ ends of reads, rather than being equally distributed along the length of the read. Using error distribution modeled on real data, we see fewer contigs reconstructed at each p.p.m., due to fewer seeds being initially identified and more frequent seed extension failures (Table 4). However, at least for clonotype frequencies $>5$ p.p.m., reconstruction success rate is high and largely unaffected by error distribution. At lower clonotype frequencies, the reconstruction rate is lower; for instance, 63.8% versus 51.5% of 1 p.p.m. sTCRβ can be identified when using a random versus a modeled error distribution, respectively (Table 4).

A change of only 1% in the read base error (from 1% to 2%) yields a 66% increase in contigs too short to characterize sTCRβ unambiguously, usually because the J segment is incomplete and/or its position cannot be identified with certainty. This translates into a decreased sensitivity of the method of over 10% at 1 p.p.m.,

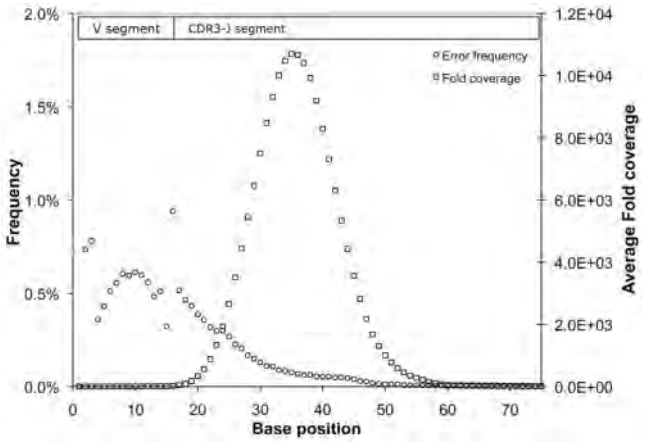



**Fig. 3.** Average mismatched base frequency and mean contig base coverage per position on trimmed, normalized, unambiguous contigs from triplicate 36 nt 1% random error read assemblies. Assembly base mismatch frequency (i.e. assembly errors) reaches a maximum within the seed portion (any of the last 15 V bases and at least eight consecutive mismatched bases downstream) and decreases through J as the base coverage increases. The sudden increase in average fold coverage beginning at approximately base position 30 is explained by over sampling of a limited number (14) of J segments, and the re-use of reads by the iSSAKE assembly algorithm. The position of the V segment and approximate position of the CDR3 and J gene segments on the contigs is shown on top of the graph and is depicted by the rectangles. Every contig is comprised of the last and first 15 nt of the V and J gene segment, respectively. The CDR3 and J gene segment boundaries are approximate because the length of the CDR3 varies.

7% at 2 p.p.m. and 2% at 3 p.p.m. At higher clonotype frequencies (>4 p.p.m.), the effect of base error on yielding short contigs is offset by the larger read depth (Table 5).

The robustness of assemblies of higher frequency clonotypes is further enhanced because more seeds are available at the start of assembly. When assembled, these seeds should, in theory, lead to contigs that characterize the same TCRβ. Keeping track of the average number of contigs that characterize each sTCRβ generated allows one to estimate the frequency of any given sTCRβ in the sample. Consistently, at the error rates tested, there is an almost perfect PEARSON correlation (0.9998, 0.9997 and 0.9994 at 1, 1.5 and 2% random error, respectively, and 0.9995 for the 1% modeled error set) between the average number of sTCRβ-capturing contigs and the frequency of that sTCRβ. Since the number of seeds (and thus, contigs) identified per TCRβ varies linearly in function of read coverage, as opposed to having a 1:1 relationship with the clonotype frequency, the number of contigs identified cannot be expected to reflect the exact clonality of each TCRβ in the sample. Instead, the contig count may be used to estimate relative TCRβ abundance.

To explore the effect of read length, we also simulated sets of 42 and 50 nt × 20M reads at 1% random error. These read lengths and error rates should be achievable by massively parallel short read platforms, if not currently, then in the near future. Increasing the read length has a drastic effect on the sensitivity of the method. Detection of sTCRβ increases by 18% at 1 p.p.m. when 42 nt 1% error reads are assembled compared to shorter 36 nt reads (Table 6). Using 50 nt reads for assembly recovers over 94.7% of clonotypes, an increase of >30% in detection compared to the usage of 36 nt reads with the same error rate. At 2 p.p.m., the sensitivity of the recovery increases from 91.1% to 97.4% to 99.1%, using 1% error 36, 42 and 50 nt reads, respectively. Increased sensitivity is a direct consequence of obtaining more seeds. With 42 and 50 nt reads, 40% and 77% more seed sequences could be identified from our set of 20M simulated reads (Table 3).

## 4 DISCUSSION AND CONCLUSION

Technological advances in sequencing (Holt and Jones, 2008) put large-scale high-resolution TCR profiling within the realm of possibility. However, shortcomings of these new sequencing technologies, namely the appreciable sequencing errors and short read lengths, require computational solutions to help make sense of the data. We have explored the feasibility of using short 36, 42 and 50 nt error-prone sequences to characterize up to 1 million sTCRβ sequences. Our strategy for reconstructing sTCRβ relies on two bioinformatics pillars: short read sequence alignment and seeded *de novo* assembly. Unidirectional *de novo* assemblies of short seeds targeting the V–D–J junction is made possible using a modified version of SSAKE (Warren *et al*., 2007) that handles sequencing errors, re-use reads and processes *k*-mers more rapidly than earlier versions (http://www.bcgsc.ca/platform/bioinfo/software/ssake). This strategy is tailored for very short reads, such as those produced by the Illumina Ltd. sequencing instrument, and constitute the main theoretical advance presented in this article.

Sequence characterization of TCRs and more specifically the variable portion encoding amino acids that directly interact with antigenic peptide permits the identification of disease-associated T-cells. Current TCR sequence profiling can at best decipher hundreds of TCRs (Ozawa *et al*., 2008; Zhou *et al*., 2006), a small number in comparison with the $10^7$ TCR diversity estimated in an individual (Arstila *et al*., 1999). Larger-scale profiling techniques that examine CDR3 length heterogeneity provide a global snapshot of TCR repertoires, but do not resolve individual clonotypes at the macromolecular level (Gorski *et al*., 1994; Pannetier *et al*., 1993; Penitente *et al*., 2008). Due to the low throughput, high cost and labor requirements of traditional Sanger sequencing, sequence-profiling TCR on that same scale has not yet been explored, thereby providing the impetus for our study.

The success of TCR sequence reconstruction using short sequences relies on the very region that makes profiling the TCR repertoire challenging; the uniqueness and specificity of the CDR3 (Davis *et al*., 1998). Selection of seed sequences that comprise bases encoding a portion of the variable region ensures that a streamlined, unidirectional assembly proceeding through the junction will help characterize unique clones. This is especially true if the sequence coverage is 10-fold or above, or the frequency of the TCR is high, since higher frequencies result in higher sequence coverage of discrete TCRs. At low frequencies, base error has a strong negative impact on TCR reconstruction rates that is due to sequence coverage insufficient to offset base error in less redundant CDR3-encoding regions. iSSAKE will not extend a seed or contig with a base error in a minimum set overlap region, unless that base can be found at the same position in an overlapping *k*-mer. This impacts favorably on contig accuracy at the expense of reconstruction rates, especially at low 1 p.p.m. frequencies and 2% error.

We examined instances of failure to detect CDR3 sequences known to exist in our sTCRβ library. The majority (99.4% using

36 nt read sets) of irresolvable low-frequency sTCRβ are attributable to 3′ base errors. The problem is exacerbated for low-frequency clones because these by definition have lower coverage and therefore less chance for an error to be mitigated by read redundancy. At all clonotype frequencies, but more noticeably at higher sequence coverage, 0.3% of irresolvable sTCRβ are due to high sequence identity between modeled TCRβ, sometimes differing only by a few 5′ V bases and thereby preventing unambiguous characterization of their sequence. We see additional failure modes that are very rare (∼0.2–0.3% of irresolvable sTCRβ, 0.004% of total sTCRβ) and associated with shorter (36 and 42 nt) reads and higher frequency clonotypes (>5 p.p.m.). For example, reads originating from CDR3 may by chance align well to V segments, C segments or JC junctions, and will as a consequence be removed early in the assembly process (Fig. 2) such that CDR3's containing these sequences cannot be assembled despite high read coverage. This failure mode was not observed with longer 50 nt seeds and is explained by the increased ability of seeds (which are never discarded) to span the entire CDR3 and capture a portion of J in a single read.

In time, accurate sequence length from next generation sequencing platforms will exceed the length of CDR3. However, sequence assembly will remain advantageous because it mitigates the effect of sequence errors present in individual reads. In the present study, fewer CDR3 sequences could be identified unambiguously with the unassembled 50 nt read set compared to the assembled one (533 868 versus 2 428 086 unambiguous contigs, Supplementary Table 1). Already, a typical Illumina Sequence Analyzer run will output more bases than we generated in this study, at equal or lower error rates. This suggests that the overall strategy, as outlined, will work with sequence data from biological samples. We expect it will also be applicable to similar metagenomics projects including sequence-characterization of the immunoglobulin repertoire.

## ACKNOWLEDGEMENTS

We thank Dr John Webb for advice. R.A.H. is a Michael Smith Foundation for Health Research Scholar.

*Funding*: Genome Canada and Genome British Columbia.

*Conflict of Interest*: none declared.

## REFERENCES

Arstila,P.T. *et al*. (1999) A direct estimate of the human αβ T cell receptor diversity. *Science*, **286**, 958–961.

Bennett,S. (2004) Solexa Ltd. *Pharmacogenomics*, **5**, 433–438.

Davis,M.M. *et al*. (1998) Ligand recognition by alpha beta T cell receptors. *Annu. Rev. Immunol.*, **16**, 523–544.

Dohm,J.C. *et al*. (2007) SHARCGS, a fast and highly accurate short-read assembly algorithm for de novo genomic sequencing. *Genome Res.*, **17**, 1697–16706.

Flicek,P. *et al*. (2008) Ensembl 2008. *Nucleic Acids Res.*, **36**, D707–D714.

Gascoigne,N.R. *et al*. (1984) Genomic organization and sequence of T-cell receptor beta-chain constant- and joining-region genes. *Nature*, **310**, 387–391.

Gorski,J. *et al*. (1994) Circulating T cell repertoire complexity in normal individuals and bone marrow recipients analyzed by CDR3 size spectratyping. Correlation with immune status. *J. Immunol.*, **152**, 5109–5119.

Hedrick,S. *et al*. (1984) Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature*, **308**, 149–153.

Heng,L. *et al*. (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.*, **18**, 1851–1858.

Holt,R.A. and Jones,S.J. (2008) The new paradigm of flow cell sequencing. *Genome Res.*, **18**, 839–846.

Huck,S. *et al*. (1988) Variable region genes in the human T-cell rearranging gamma (TRG) locus: V-J junction and homology with the mouse genes. *EMBO J.*, **7**, 719–726.

Janeway,C.A. Jr *et al*. (2001) *Immunobiology*. 6th edn. Garland Science, London, UK.

Jeck,W.R. *et al*. (2007) Extending assembly of short DNA sequences to handle error. *Bioinformatics*, **23**, 2942–2944.

Kavaler,J. *et al*. (1984) Localization of a T-cell receptor diversity-region element. *Nature*, **310**, 421–423.

Landau,N.R. *et al*. (1984) Cloning of terminal transferase cDNA by antibody screening. *Proc. Natl Acad. Sci. USA*, **81**, 5836–5840.

Lefranc,M.-P. and Lefranc,G. (2001) *The T cell Receptor Facts-Book*. Academic Press, London, UK.

Margulies,M. *et al*. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.

Ozawa,T. *et al*. (2008) Comprehensive analysis of the functional TCR repertoire at the single-cell level. *Biochem. Biophys. Res. Commun.*, **367**, 820–825.

Pannetier,C. *et al*. (1993) The sizes of the CDR3 hypervariable regions of the murine T-cell receptor beta chains vary as a function of the recombined germ-line segments. *Proc. Natl Acad. Sci. USA*, **90**, 4319–4323.

Penitente,R. *et al*. (2008) Administration of PLP139-151 primes T cells distinct from those spontaneously responsive in vitro to this antigen. *J. Immunol.*, **180**, 6611–6622.

Sakano,H. *et al*. (1979) Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature*, **280**, 288–294.

Slater,G.S. and Birney,E. (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, **6**, 31.

Warren,R.L. *et al*. (2007) Assembling millions of short DNA sequences using SSAKE. *Bioinformatics*, **23**, 500–501.

Zhou,D. *et al*. (2006) High throughput analysis of TCR-b rearrangement and gene expression in single T cells. *Lab. Invest.*, **86**, 314–321.

# Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing

J. Douglas Freeman,[1,3] René L. Warren,[1,3] John R. Webb,[2] Brad H. Nelson,[2] and Robert A. Holt[1,4]

[1]*BC Cancer Agency, Michael Smith Genome Sciences Centre, Vancouver, British Columbia V5Z 1L3, Canada;* [2]*BC Cancer Agency, Deeley Research Centre, Victoria, British Columbia V8R 6V5, Canada*

**T-cell receptor (TCR) genomic loci undergo somatic V(D)J recombination, plus the addition/subtraction of nontemplated bases at recombination junctions, in order to generate the repertoire of structurally diverse T cells necessary for antigen recognition. TCR beta subunits can be unambiguously identified by their hypervariable CDR3 (Complement Determining Region 3) sequence. This is the site of V(D)J recombination encoding the principal site of antigen contact. The complexity and dynamics of the T-cell repertoire remain unknown because the potential repertoire size has made conventional sequence analysis intractable. Here, we use 5′-RACE, Illumina sequencing, and a novel short read assembly strategy to sample CDR3$_\beta$ diversity in human T lymphocytes from peripheral blood. Assembly of 40.5 million short reads identified 33,664 distinct TCR$_\beta$ clonotypes and provides precise measurements of CDR3$_\beta$ length diversity, usage of nontemplated bases, sequence convergence, and preferences for *TRBV* (T-cell receptor beta variable gene) and *TRBJ* (T-cell receptor beta joining gene) gene usage and pairing. CDR3 length between conserved residues of *TRBV* and *TRBJ* ranged from 21 to 81 nucleotides (nt). *TRBV* gene usage ranged from 0.01% for *TRBV17* to 24.6% for *TRBV20–1*. *TRBJ* gene usage ranged from 1.6% for *TRBJ2–6* to 17.2% for *TRBJ2–1*. We identified 1573 examples of convergence where the same amino acid translation was specified by distinct CDR3$_\beta$ nucleotide sequences. Direct sequence-based immunoprofiling will likely prove to be a useful tool for understanding repertoire dynamics in response to immune challenge, without a priori knowledge of antigen.**

**[Supplemental material is available online at http://www.genome.org. The TCR$_\beta$ cDNA sequence and quality-score files have been submitted to the NCBI Short Read Archive (http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi) under accession no. SRA008633.]**

T-cell receptors (TCRs) are dimeric ($\alpha\beta$ or $\gamma\delta$), highly variable T lymphocyte membrane proteins that recognize antigenic peptides presented on heterologous cells by the major histocompatibility complex (MHC) (Davis and Bjorkman 1988; Bassing et al. 2002). Recognition specificity for diverse peptide-MHC (pMHC) complexes is provided by the three complementarity-determining regions (CDRs) of the TCR. CDR1 and CDR2 are coded for by germline sequences while CDR3, the highly polymorphic principal recognition site, is created when TCR genomic loci undergo somatic recombination between gene segments during development of T lymphocytes in the thymus (Gellert 1992, 2002; Jung and Alt 2004). For the $\alpha$ locus and the $\gamma$ locus, recombination occurs between variable (V) and joining (J) segments. For the $\delta$ locus and the $\beta$ locus, there is recombination between V and J segments, but also the inclusion of one of two short diversity (D) segments. The combinatorial diversity of the human $\beta$ locus is illustrated in Figure 1A. At CDR3 recombination junctions, further complexity is generated through the deletion of germline-encoded bases and the addition of random nontemplated bases. The resulting hypervariable sequences of the CDR3 make possible the recognition of diverse peptide-MHC (pMHC) complexes. During T-cell maturation, all T cells expressing rearranged receptors capable of binding pMHC with high enough affinity to be biologically relevant are retained (positive selection), but only T cells with rearranged receptors that do not interact strongly with self-pMHC complexes ultimately exit the thymus (negative selection). It should be noted that V(D)J recombination is not entirely random, and the prevalence of specific gene segments and combinations of gene segments shows marked variation in the repertoire. Contributions to this bias are introduced even before thymic selection, through variation in the efficiency of recombination of different gene segments (Manfras et al. 1999; Krangel 2003). The peripheral blood thus contains a large repertoire of T lymphocytes with the potential to recognize diverse antigens. Binding of a naïve T cell's TCR to a structurally compatible pMHC on an antigen-presenting cell will, with the appropriate interaction of co-stimulatory molecules, initiate rapid clonal expansion to generate a population of effector cells. This acute response occurs on the order of days, and is followed by a gradual contraction of the expanded pool over the course of several weeks, with differentiation into a small number of long-lived memory cells. Thus, the T-cell repertoire is not static, but rather is constantly molded by immune challenge (for reviews, see Nikolich-Zugich et al. 2004; Harty and Badovinac 2008).

There has been remarkable progress in characterizing the size and dynamics of the T-cell repertoire, but the task remains daunting due to the enormous combinatorial diversity that is theoretically possible (>$10^{15}$ distinct $\alpha\beta$ receptors, or clonotypes [Davis and Bjorkman 1988; Murphy et al. 2007]) and the limited power of existing tools for interrogation. Previously, a method called TCR spectratyping (Pannetier et al. 1993; Gorski et al. 1994) had been used to probe the T-cell repertoire. This approach involves the use of V and J gene segment-specific primers for RT-PCR amplification of the CDR3. In TCR spectratyping, CDR3 amplicons are separated according to size by polyacrylamide gel electrophoresis. Typically, six or so distinct amplicons are observed per primer pair, spaced at 3-nucleotide (nt) intervals in accordance

**[3]These authors contributed equally to this work.**
**[4]Corresponding author.**
**E-mail rholt@bcgsc.ca; fax (604) 877-6085.**
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.092924.109.

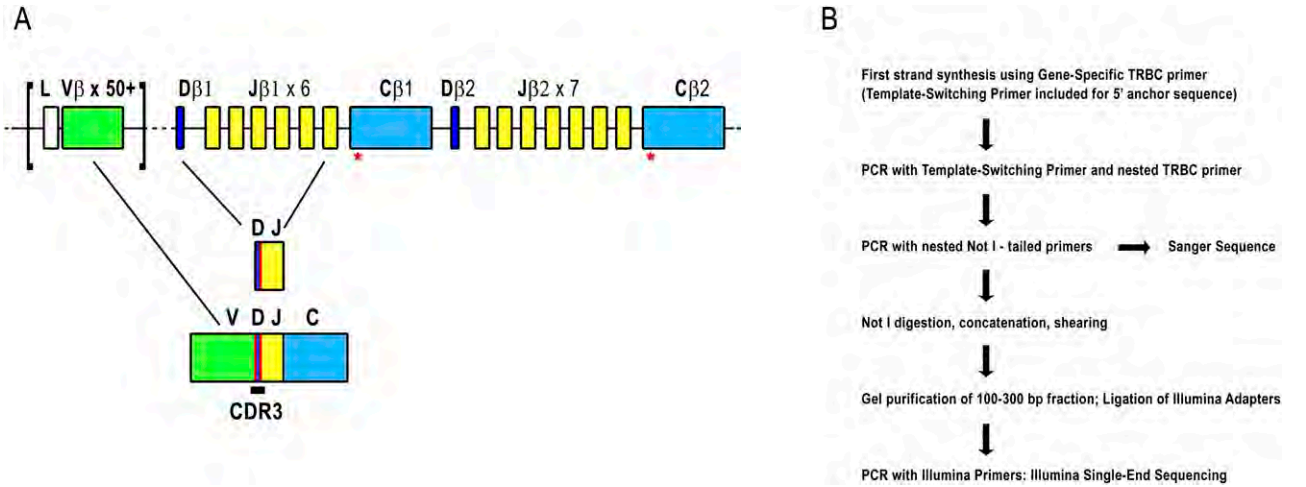



**Figure 1.** (*A*) Representation of the $TCR_\beta$ locus at human chromosome 7q34. The $TCR_\beta$ locus spans 620 kb and includes over 50 *TRBV* genes (green) belonging to 30 subgroups. There are two *TRBC* genes (light blue) each downstream from a *TRBD* (dark blue) and six or seven *TRBJs* (yellow). Recombination first occurs between *TRBJ* and *TRBD* genes, followed by recombination to a *TRBV* gene. (Red lines) Addition of nontemplated bases. After transcription, intervening sequences are spliced out so that a *TRBC* is adjacent to the recombined V-D-J sequence. Gene width and distances are not to scale. (Red asterisks under *TRBC*) Location of primers used for 5′-RACE and PCR reactions. Refer to Supplemental Figure 1A for primer locations. A detailed locus map can be obtained from IMGT (www.imgt.org/textes/IMGTrepertoire). (*B*) Flowchart illustrating 5′-RACE and Illumina library construction. For more detail, please refer to Methods and Supplemental Figure 1.

with reading frame. An experimental estimate of repertoire size of $\sim 10^6$ beta chains in blood has been obtained (Arstila et al. 1999) by exhaustive Sanger sequencing of a single amplicon from a *TRBV18-TRBJ1-4* spectratype, then extrapolating the observed diversity according to the relative abundance of this amplicon in the spectratype and the estimated frequency of *TRBV18-TRBJ1-4* pairing in the repertoire. Of course, actual TCR diversity will be higher still, due to αβ heterodimerization (Fuschiotti et al. 2007; Ozawa et al. 2008).

Advances in sequencing technology (Holt and Jones 2008; Shendure and Ji 2008) now permit interrogation of complex sequencing targets at unprecedented depth and reasonable cost. Here, we describe a method for deep sampling of the TCR repertoire at sequence-level resolution. Our approach relies on massively parallel Illumina sequencing of $CDR3_\beta$ amplification products and a novel TCR-specific short read assembly strategy (Warren et al. 2009).

## Results

### Experimental strategy

We used 5′ rapid amplification of cDNA ends (RACE) to obtain $CDR3_\beta$ transcript sequences from a commercially available mRNA sample prepared from normal human peripheral blood leukocytes (PBL) pooled from 550 individuals (Fig. 1B; Supplemental Fig. 1). Peripheral blood from different individuals will include different frequencies of naïve and memory T cells. Because individual memory repertoires are skewed due to historical antigen encounter and the individual's HLA type, our results do not reflect the expected repertoire of any individual, but rather are reflective of average clonotype abundance in a population.

The RACE approach avoids the potential bias associated with the use of the multiple primer sets required to amplify from all TRBV sequences (Boria et al. 2008) and takes advantage of the conserved sequences offered by *TRBC1* and *TRBC2* (96% nucleotide sequence identity). Reverse transcription to generate cDNA was performed using a primer specific for the *TRBC* genes (Ozawa et al. 2008) as well as a template-switching primer (Peters et al. 1999; Douek et al. 2002) to provide a 5′ anchor for subsequent PCR. First-round PCR reactions with a nested *TRBC* primer and the template-switching primer produced a high level of background amplification. A second round of PCR using nested primers was performed to obtain a cleaner product of ~520 bp. (See Methods for primer sequences and Supplemental Fig. 1A for *TRBC* primer locations.) The RACE product was then gel-purified and an aliquot was cloned and Sanger sequenced to confirm the presence of $CDR3_\beta$ amplicons. The RACE product was too long to directly sequence the $CDR3_\beta$ region with short-read technology, so it was ligated to produce concatamers that were then sheared by sonication. A 100- to 300-bp size fraction was isolated by PAGE and shotgun-sequenced on the Illumina platform (www.illumina.com). The initial sequencing runs generated 18,829,563 36-nt reads. During the course of this analysis, a protocol to produce longer read lengths became available, so further 21,752,666 50-nt reads were generated and analysis was performed on the pooled set of 40,582,229 reads (Table 1).

### iSSAKE assembly and analysis of reconstructed $TCR_\beta$ sequences

We have recently described a system for profiling TCR diversity using short sequence reads and the assembly software package we call iSSAKE (immuno-Short Sequence Assembly by *K*-mer search

**Table 1. Sequencing and assembly statistics**

| | |
|---|---|
| Total reads | 40,582,229 |
| Seed sequences | 310,614 |
| Total $CDR3_\beta$ sequences assembled[a] | 117,052 |
| Total clonotypes (distinct $CDR3_\beta$ sequences) | 33,664 |
| Clonotypes with an unambiguous *TRBV* segment | 22,704 |

[a]Complete $CDR3_\beta$ sequences in correct reading frame.

and 3′ read Extension; Warren et al. 2009). Sequence reads that have homology with a *TRBV* gene segment but have unmatched bases at their ends (corresponding to the beginning of the recombined $CDR3_\beta$ sequence) are used as seeds to initiate directional, de novo $CDR3_\beta$ assemblies, as described in the Methods section. Briefly, with iSSAKE, reads from the shotgun data set derived from the $CDR3_\beta$ amplicon are optimally aligned to each seed to generate $CDR3_\beta$ sequence contigs. An important feature of iSSAKE is that reads can be reused, which allows the assembly of different $CDR3_\beta$ sequences that end in the same *TRBJ* segment. However, this also means that read depth is not uniform throughout $CDR3_\beta$ and cannot be used to determine clonotype frequency. Depth is exceptionally high over the *TRBJ* segment, where only 14 *TRBJ* possibilities exist. Some seeds are themselves long enough to cover most or all of the $CDR3_\beta$-encoded bases without being significantly extended by other reads, so read depth in these instances may be low. For the current study, we set the assembly parameters to only extend a contig when two or more reads aligned at each base position (−o 2). Further, only reads with overlaps 15 bases or longer (−m 15) were considered. Read error was mitigated using the iSSAKE internal error-handling algorithm and consensus bases were called when bases from reads agreed 70% of the time or more (−r 0.7). The output comprises contiguous sequences (contigs) that contain the last 15 nt of *TRBV*, any nontemplated and *TRBD* bases, and the first 15 recognizable *TRBJ* segment bases. Upon completion of assembly, contigs are compared and those that have matching sequence, and therefore represent the same beta-chain clonotype, are grouped together. The number of matching contigs for a given clonotype provides the relative frequency of that clonotype in the original sample. From previous simulations we know that contig depth determined in this manner is proportional to clonotype frequency (correlation coefficient >0.999; Warren et al. 2009).

The results of the analysis of the sequence data generated in the present study are outlined in Table 1. From the complete data set of >40.5 M total reads we identified 310,614 assembly seeds. From these seeds, $CDR3_\beta$ sequences were assembled for 35,762 distinct TCR beta-chain clonotypes. The sequence of each clonotype is provided in Supplemental File 1. A large proportion of seeds did not yield additional distinct clonotypes, as there was not adequate sequence coverage of our sample to extend these seeds into *TRBJ*. Clonotype sequences were screened for open reading frames (ORFs), and 2098 (5.9%) that were found to contain stop codons in frame with *TRBJ* were removed, leaving 33,664 distinct beta-chain clonotypes. The relatively high proportion of clonotypes containing stop codons does not appear to be due to sequencing error or misassembly. Rather, we find that 96% of these stop codons map to nontemplated bases at the V-D-J junction and likely represent real events captured by our assay. During T-cell maturation, in the event that the first T-cell receptor beta chain ($TCR_\beta$) rearrangement in a given cell is nonproductive, rearrangement of the second $TCR_\beta$ allele is initiated. However, after thymic selection, downregulation of the nonproductive allele is not absolute (Li and Wilkinson 1998), and these transcripts may account for the premature termination codons we identify in the present study.

Clonotype abundance varied from one to a maximum of 279 (Fig. 2A). Rare clonotypes detected as single copies represented 65.3% of all clonotypes but only 18.8% of the 117,052 total $CDR3_\beta$ sequence assemblies. Moderately abundant clonotypes detected at a copy number between two and 19 represented 32.6% of clonotypes and 64.1% of assemblies. Finally, there were 720 clonotypes (2.1% of clonotypes) with copy number >20, and this small number of highly abundant clonotypes represented 17.1% of all assemblies. Since the data are generated from RNA pooled from many individuals, we expect that the majority of clonotypes will originate from the more prevalent effector and memory cells of the population sampled. It is possible that the most abundant clonotypes therefore represent highly expanded effector cells from those individuals with a recent antigen encounter. Additionally, some abundant clonotypes may exemplify the phenomenon of public T-cell responses, that is, identical TCR rearrangements from multiple individuals in response to the same antigen (Venturi et al. 2008).

To determine if the depth of sequencing in the current study showed any trend toward saturation, we took random subsamples of sequences at intervals of 5 million reads. These were independently assembled and the number of distinct clonotypes at each point was plotted (Fig. 2B). The relationship is linear (Pearson coefficient = 0.999), indicating we have not begun to approach saturation. This is expected, given the fact that we obtained 33,664 distinct clonotypes from our sequencing target of pooled

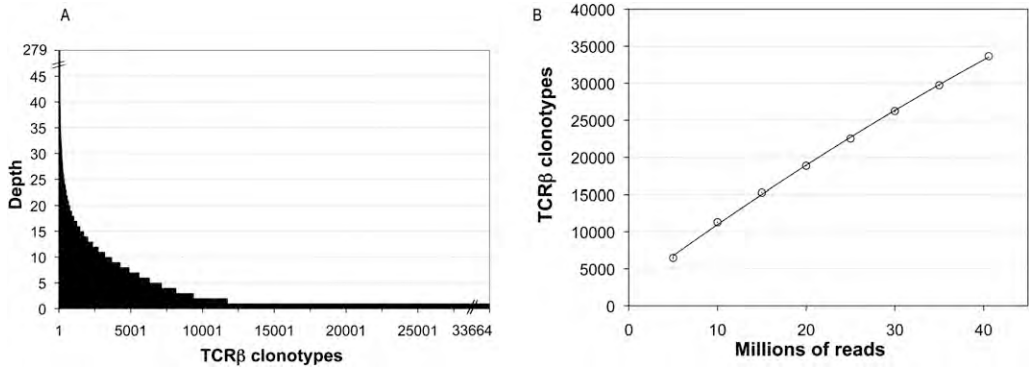



**Figure 2.** (*A*) $TCR_\beta$ diversity. A total of 33,664 $TCR_\beta$ clonotypes were identified from complete and in-frame $CDR3_\beta$ sequences assembled by iSSAKE. Clonotypes with a copy number of one (clonotypes identified by a single iSSAKE contig) account for 65.3% of all clonotypes. Clonotypes identified by two to 19 iSSAKE contigs account for 32.6% of all clonotypes, and high-abundance clonotypes (contig depth ≥20) account for 2.1% of the total. (*B*) Saturation analysis. In duplicate experiments we chose independent sets of 5, 10, 15, 20, 30, and 35 M reads at random from the pool of 40,582,229 total sequence reads in our data set. These subsets of reads were assembled and clonotypes counted as the set of complete, in-frame, nonredundant $CDR3_\beta$ sequences. The number of clonotypes (mean ± SD for the duplicate experiments) is plotted as a function of the number of reads. Error bars are contained *within* the symbols.

peripheral blood mononuclear cells (PBMCs), and, as noted above, the repertoire size of just one individual has been estimated previously to be $\sim 10^6$ beta chains.

### *TRBV*, *TRBD*, and *TRBJ* usage

With our experimental approach, only the portion of the *TRBV* sequence that is present in the seed sequence is available for assignment of a particular *TRBV* gene to an assembled CDR3$_\beta$. This fact, together with the often high sequence homology among certain *TRBV* genes and the replacement of deleted *TRBV* ends with nontemplated bases makes it impossible to assign a single unique *TRBV* gene to every clonotype. We were successful, however, in making an unambiguous assignment for 22,704 (67.4%) of the assembled clonotypes (Table 1; Supplemental Fig. 2) to one of 49 distinct *TBRV* genes. Subsequent analysis is based on this portion of the data set. Usage of this set of 49 *TBRV* genes ranged from 24.6% for *TRBV20-1* to 0.01% for *TRBV17* (Fig. 3A).

Several *TRBV* genes identified in our assembly are ORFs (open reading frame, International ImMunoGeneTics [IMGT] nomenclature) that either have changes at conserved amino acid positions (*TRBV6-7*, *TRBV17*, and *TRBV7-1)* or noncanonical splice donor sites (*TRBV5-3* and *TRBV23-1*). We also found assemblies containing the pseudogene *TRBV21-1*, which has a frameshift in the leader sequence. As with the appearance of rearrangements that contain stop codons (see above), it is possible that these assemblies represent the transcription of an initial nonproductive rearrangement. *TRBV21-1* and *TRBV23-1* have previously been shown to be transcribed (summarized by Folch and Lefranc 2000), and for *TRBV23-1*, expression has also been demonstrated at the protein level (Leslie et al. 2006). The transcription of *TRBV7-1* is unexpected, as it is also deficient in sequences essential for recombination. It is possible that functional alleles of *TRBV7-1* exist.

All known, functional *TRBJ* genes are represented in the sequence assembly and can be assigned unambiguously within the full set of 33,664 potential clonotype sequences. Usage ranged from 17.2% for *TRBJ2-1* to 1.6% for *TRBJ2-6* (Fig. 3B). The pseudogene *TRBJ2-2P* was not detected.

*TRBD* segments sustain substantial base deletion and overall transformation, so that the segments are usually unrecognizable without ambiguity. Clonotypes where *TRBD* could be identified unambiguously and accurately (with minimum length = 8 nt) represent 5497 out of the 22,704 sequences with accurate *TRBV* assignment. For these 5597 clonotypes, we calculate that 49.9% are *TRBD1* and 50.1% are *TRBD2*.

From our set of 49 positively identified *TRBV* and 13 *TRBJ*, there are 637 potential pairings. We find that 562 of these *TRBV–TRBJ* combinations are represented in our data for the 22,704 unique clonotypes (Fig. 4; Supplemental Fig. 3; Supplemental Table 1). The most frequent pairing is of *TRBV20-1* to *TRBJ2-1*, accounting for 4.1% of all pairings, whereas 58 *TRBV–TRBJ* pairs were identified only once.

### Sequence diversity of CDR3$_\beta$

We examined the frequency of base addition and deletion at the V-D-J junction. The boundary of the CDR3$_\beta$ is not absolute. In order to compare sequences, we defined CDR3$_\beta$ coordinates as starting at the codon for the last cysteine of *TRBV* and ending at the phenylalanine in the conserved *TRBJ* segment motif FGXG. In our data set this defines a subset of 30,366 CDR3$_\beta$ sequences. The length of CDR3 varies from 21 to 81 nt with a peak at 45 nt (Fig. 5A). While most rearrangements involve the removal and addition of a few residues, the extent of change can be considerable. Nontemplated bases and/or deletions were detected in all D-J junctions examined and in only two instances was there no net change in sequence at the V-D junction. Nontemplated bases at the V-D junction are 62.9% GC and the nontemplated bases at the J-D junction are 54.3% GC. The 7319 CDR3$_\beta$ sequences contained in the 45-nt peak were used to create nucleotide and amino acid sequence logos (Fig. 5B,C). The logos are a graphical representation of a nucleic acid or amino acid multiple sequence alignment (Schneider and Stephens 1990; Crooks et al. 2004). We find no evidence of any overrepresented sequence in CDR3$_\beta$ other than the prevalence of guanines in the center of the nucleic acid logo (Fig. 5B) and glycines in the center of the amino acid logo (Fig. 5C), which simply reflect the sequence and coding potential of the guanine-rich *TRBD* segments. The conservation apparent at the left and right ends of the logos reflects the contribution of *TRBV* and *TRBJ* sequences, respectively. Finally, in our sequence assemblies, there are many examples of independent recombination events that have produced the same CDR3$_\beta$ amino acid sequence. In 659 instances, the same CDR3$_\beta$ nucleotide sequence is detected in association with different *TRBV* and *TRBJ* sequences. In addition, we find 1573 examples of rearrangements where the same CDR3$_\beta$ amino acid sequence can be translated from divergent nucleotide sequences.

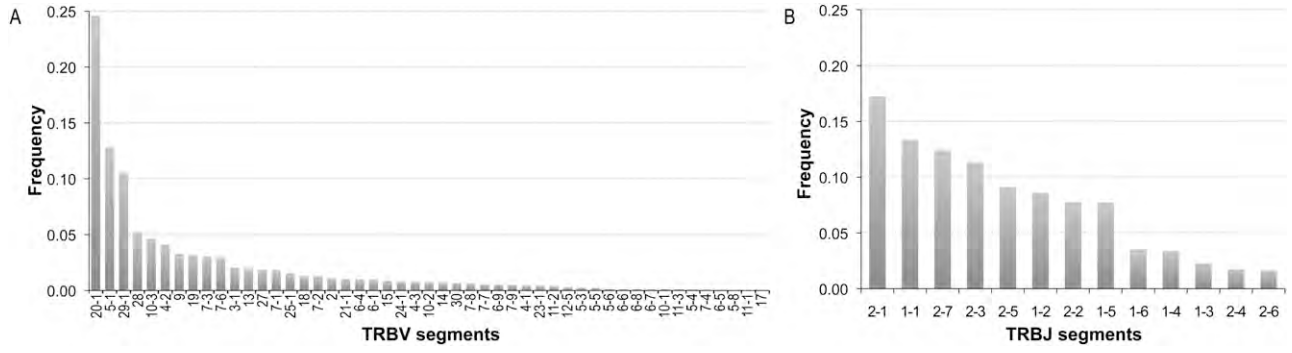



**Figure 3.** *TRBV* and *TRBJ* usage. (*A*) Relative frequency of usage of *TRBV* segments was for the subset of clonotypes with an unambiguous *TRBV* gene segment assignment ($n$ = 22,704). (*B*) Relative frequency of usage of *TRBJ* segments for the set of all clonotypes ($n$ = 33,664).

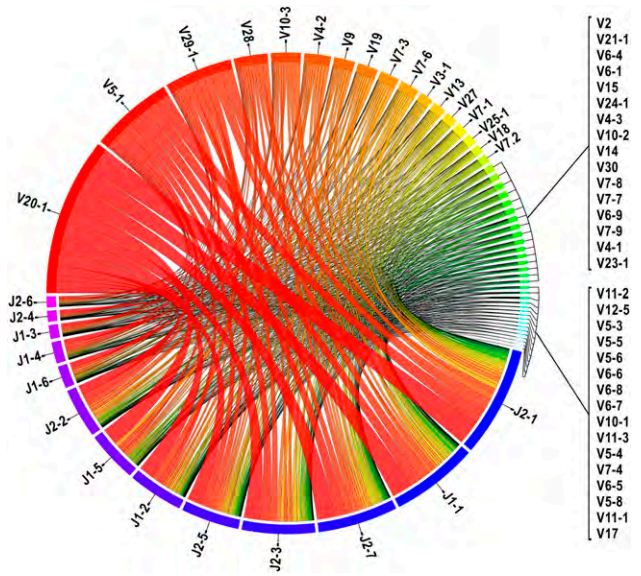



**Figure 4.** Frequencies of V-J pairing calculated from the subset of clonotypes with an unambiguous *TRBV* gene segment assignment ($n$ = 22,704). (Blue to purple rectangular bands) *TRBJ* segments, (red to cyan rectangular bands) *TRBV* segments. The width of the bands is proportional to the number of times the *TRBV* and *TRBJ* genes connected by the band co-occur in CDR3$_\beta$ sequences. *TRBV* and *TRBJ* segments are arranged from *left* to *right* and *right* to *left*, respectively, and ordered by total pairing links they share. (This illustrates the data contained in Supplemental Table 1.) This figure was generated using the Circos software package (Krzywinski et al. 2009).

## Verification

Validation of the assembly and analysis was obtained from the intersection of iSSAKE-generated data with that obtained by Sanger sequencing of 5′-RACE clones. A total of 288 unique TCR sequences were obtained by Sanger sequencing of the cloned 5′-RACE products, and, of these, 69% were found in our assembled Illumina data. In addition, comparison of the usage frequencies of *TRBJ* genes from the iSSAKE contigs and the Sanger sequences shows a very similar profile (Pearson coefficient 0.904, data not shown) and provides a validation of the iSSAKE sequence assembly. Comparison of the frequencies for *TRBV* was limited to those 49 genes that could be unambiguously identified from the Illumina assembly (Pearson coefficient 0.822). Next, to estimate sequence error attributable to PCR and reverse transcription, we used Sanger sequencing to evaluate TCR$_\beta$ 5′-RACE products amplified and cloned from the cell line DES M26, a melanoma-specific T-cell line that carries a single TCR$_\beta$ rearrangement. Single-pass sequence reads from 311 independent colonies were quality-trimmed to Q50 (representing the highest quality subset with only one predicted error per 100,000 bases) and assembled, demonstrating an accuracy of 99.91%. This sets a practical upper limit on the contribution of reverse transcriptase- and polymerase-introduced errors in the method as it is described here.

## Discussion

By massively parallel sequencing of CDR3$_\beta$ amplicons from pooled leukocytes, we have identified 33,664 distinct human CDR3$_\beta$ sequences. As of May 2009 there were only 3187 unique human TCR$_\beta$ mRNA sequences in GenBank, not all of which include the CDR3$_\beta$ region. The IMGT database reports 5303 rearranged human TCR$_\beta$ sequences, 2927 of which are shared with GenBank. Thus, in a single experiment, we have increased the number of known sequences by an order of magnitude. Analysis of the data from the present study has provided information, at unprecedented precision, on many fundamental TCR$_\beta$ properties such as preferences for nucleotide removal and addition at recombination junctions (Table 2) and the extent of CDR3$_\beta$ length diversity (Fig. 5A). As expected from all previous studies, we see that certain *TRBV* and *TRBJ* genes are commonly utilized while others are quite rare (Fig. 3A,B) and the pairing of *TRBV* and *TRBJ* is not random (Fig. 4; Supplemental Table 1; Rosenberg et al. 1992; Even et al. 1995; Hall and Lanchbury 1995; Roldan et al. 1995; Manfras et al. 1999). The reasons for bias are not clearly understood but are likely due to a combination of proximity effects and recombination signal sequence compatibilities that influence initial TCR development, plus thymic selection and immune challenge that modify the representation of selected clones in the extant repertoire (Krangel 2003). It must be emphasized that the results presented here for *TRBV* and *TRBJ* frequency and pairing are obtained from a biased sample, where the individual repertoires of subjects contributing to the pool have been skewed by antigen encounter and the individual's HLA type. These results cannot be taken to represent the innate *TRBV* and *TRBJ* usage and pairing preferences of the

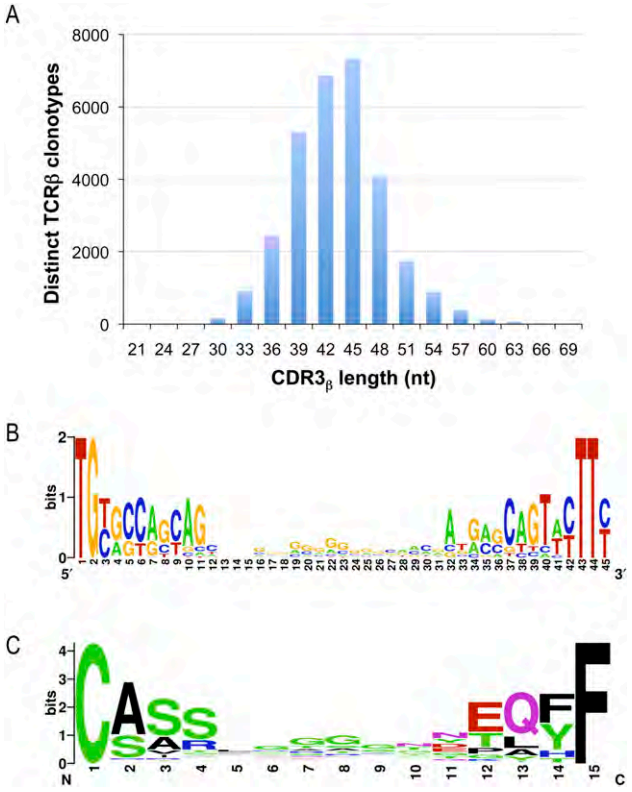



**Figure 5.** CDR3$_\beta$ nucleotide length distribution and sequence composition of the most abundant CDR3$_\beta$ length. To explore CDR3$_\beta$ length variation we used a precise length criterion, defined as all bases between the last cysteine of *TRBV* and the phenylalanine in the *TRBJ* segment motif FGXG. Of the 33,664 total clonotypes, 30,366 could be classified in this manner, and the length distribution of this subset was plotted (*A*). The most frequently observed length was 45 nt. For the subset of clonotypes with 45 nt CDR3$_\beta$ sequences, we created logos for the nucleotide (*B*) and inferred amino acid (*C*) composition, using WebLogo (Crooks et al. 2004).

**Table 2.** Frequency of base addition and deletion at the $TCR_\beta$ V-D-J junction

| Bases | Deleted 3′-V bases | Added V-D bases | Deleted 5′-D bases[a] | Deleted 3′-D bases[a] | Added D-J bases | Deleted 5′-J bases |
|---|---|---|---|---|---|---|
| 0 | 0.2384 | 0.1625 | 0.3231 | 0.2401 | 0.1246 | 0.1197 |
| 1 | 0.1629 | 0.1179 | 0.1683 | 0.1621 | 0.1035 | 0.0867 |
| 2 | 0.1008 | 0.1466 | 0.1299 | 0.2432 | 0.1457 | 0.1037 |
| 3 | 0.1159 | 0.1403 | 0.0746 | 0.2072 | 0.1466 | 0.0930 |
| 4 | 0.1359 | 0.1330 | 0.1077 | 0.0631 | 0.1257 | 0.1253 |
| 5 | 0.1044 | 0.0817 | 0.0555 | 0.0355 | 0.0975 | 0.1143 |
| 6 | 0.0743 | 0.0635 | 0.0504 | 0.0202 | 0.0793 | 0.0897 |
| 7 | 0.0391 | 0.0480 | 0.0444 | 0.0133 | 0.0389 | 0.0734 |
| 8 | 0.0152 | 0.0357 | 0.0462 | 0.0153 | 0.0369 | 0.0548 |
| 9 | 0.0061 | 0.0233 | | | 0.0227 | 0.0327 |
| 10 | 0.0043 | 0.0184 | | | 0.0133 | 0.0207 |
| 11 | 0.0022 | 0.0073 | | | 0.0164 | 0.0131 |
| 12 | 0.0004 | 0.0073 | | | 0.0076 | 0.0101 |
| 13 | $4.40 \times 10^{-05}$ | 0.0071 | | | 0.0135 | 0.0095 |
| 14 | | 0.0033 | | | 0.0053 | 0.0072 |
| 15 | | 0.0013 | | | 0.0060 | 0.0139 |
| 16 | | 0.0009 | | | 0.0045 | 0.0059 |
| 17 | | 0.0005 | | | 0.0045 | 0.0077 |
| 18 | | 0.0002 | | | 0.0029 | 0.0065 |
| 19 | | 0.0004 | | | 0.0020 | 0.0032 |
| 20 | | 0.0002 | | | 0.0009 | 0.0027 |
| 21 | | na | | | 0.0005 | 0.0019 |
| 22 | | 0.0009 | | | 0.0002 | 0.0013 |
| 23 | | | | | | 0.0017 |
| 24 | | | | | | 0.0008 |
| 25 | | | | | | 0.0005 |

[a]D segments (D1: GGGACAGGGGGC, D2: GGGACTAGCGGG[AG]GGG) were trimmed 1 bp at a time and used for the search until D=8 nt.

V(D)J recombination process during T-cell development. Further, it is possible that preferential PCR amplification of certain *TRBV* and *TRBJ* sequences over others has skewed the usage frequencies reported here. We have attempted to mitigate this by using 5′-RACE, rather than *TRBV*-specific primers, in order to reduce bias from differential primer annealing or variable amplicon lengths. The uniformity of the CDR3 length distribution presented in Figure 5 suggests that length bias has not been an issue. However, future studies with independent replicates from multiple individuals may be informative regarding other, unanticipated sources of bias.

To differentiate the bias in *TRBV* and *TRBJ* representation that is incurred during V(D)J recombination and thymic selection from bias that is cased by antigen encounter in the periphery, it should be possible to sort and independently profile the naïve ($CD25^-$, $CD44^-$, $CD45RA^+$, $CD62L^+$) and memory/effector ($CD25^+$, $CD44^+$, $CD45RO^+$, $CD62L^-$) peripheral T-cell compartments. However, biases that are strictly due to recombination can only be delineated by profiling rearranged but pre-selected double-negative (CD4−, CD8−) cells from the thymic cortex. Studies of this nature would be most tractable in mouse.

Individual TCR diversity has been estimated as $\sim 10^6$ beta chains (Arstila et al. 1999). This estimate relied on the calculation that *TRBV18* to *TRBJ1-4* pairing would occur at a frequency of 0.00024 in the repertoire. Our findings do not conflict with this assumption. We see pairing of *TRBV18* and *TRBJ1-4* in 4 of 22,704 unique clonotypes (which represents a frequency of 0.00018). However, at the present time we cannot provide insight into individual repertoire size because our sample is derived from blood pooled from multiple individuals. In fact, there may not be a "typical" individual peripheral blood T-cell repertoire, given that repertoires are skewed by many factors, including HLA type, historical antigen encounters, and current responses to acute infection.

There is considerable utility in our method of TCR sequence profiling (or its adaptation to immunoglobulin sequencing) even in the absence of heroic sequencing depth, since clonotypes that respond to a given immune challenge (effector and subsequently memory cells) will be present in the peripheral blood in higher copy number. There are many types of immune challenge, for example, acute and latent infections, autoimmunity, organ transplantation, and vaccination against infectious agents and malignancies. In these scenarios, and without a priori knowledge of the antigen, sequencing prospective samples to moderate depth should reveal responsive clonotypes. Additional applications may include the evaluation of immune reconstitution following, for example, bone marrow transplant or the initiation of highly active antiretroviral therapy (HAART) for HIV infection. It should also be possible to more readily identify the public T-cell clonotypes associated with infectious agents or tumor neoantigens and correlate these with effective immune responses.

## Methods

### 5′-RACE

Peripheral leukocyte polyA+ RNA isolated from 165 L of peripheral blood pooled from 380 males (ages 18–40) and 170 females (ages 18–40) was obtained from a commercial supplier (Clontech #636170). First-strand cDNA was synthesized using a published *TRBC* primer (5′>CACGTGGTCGGGGWAGAAGC<3′) (Ozawa et al. 2008). A target-switching oligo (5′>AAGCAGTGGTAACAACGCA GAGTACGCGGG<3′) (Peters et al. 1999) was added to provide a 5′ template for RACE (reaction conditions: 100 ng of RNA, oligonucleotides 1 μM each, 2 mM DTT, 1 mM each dNTP, 25 mM Tris-HCl pH 8.3, 37.5 mM KCl, 1.5 mM $MgCl_2$, and 400 units of Superscript II [Invitrogen] in a 20-μL volume. Extension was 90 min at 42°C followed by inactivation for 7 min at 72°C.). A control reaction with no enzyme was included to ensure that subsequent PCR products were the result of amplification from a reverse-transcribed template. PCR was performed using Platinum *Pfx* (Invitrogen) and 0.5 μL of first-strand reactions with the target-switching oligonucleotide (above) and a nested *TRBC* primer (5′> TGGTGCGGCCGCTCTCTGCTTCTGATGGCTCAAAC<3′) tailed with a NotI restriction site (reaction conditions: 1 unit of enzyme, 2× *Pfx* amplification buffer, 1 mM $MgSO_4$, oligonucleotides 0.3 μM each, and 0.3 mM each dNTP in a 50-μL volume; 2 min denaturation at 94°C was followed by 30 cycles of 30 sec at 94°C, 30 sec at 55°C, and 45 sec at 68°C, plus a final extension for 5 min at 68°C). In order to obtain a cleaner product, PCR was performed on 0.1 μL of the first-round reaction with a nested target-switching oligonucleotide (5′>AGTTGCGGCCGCTAACAACGCAGAGTACG CGGG<3′), and an equimolar combination of two primers (5′>CA CAGCGGCCGCGGGTGGGAACACCTTGTTCAGGT<3′), and (5′> CACAGCGGCCGCGGGTGGGAACACGTTTTTCAGGT<3′) specific for *TRBC1* and *TRBC2*, respectively (reaction conditions: 1 unit of

enzyme, 2× *Pfx* amplification buffer, 1 mM $MgSO_4$, 0.3 μM oligonucleotides, and 0.3 mM each dNTP in a 50-μL volume; 2 min denaturation at 94°C was followed by 20 cycles of 30 sec at 94°C and 75 sec at 68°C, plus a final extension for 5 min at 68°C.). The nested PCR reaction was loaded on a 12% polyacrylamide gel and the band centered at 520 bp was visualized using SYBR Green (Lonza), excised, and processed for sequencing (below).

### Preparation of 5′-RACE products for Illumina sequencing

Eight nested PCR reactions were pooled and purified using 20 μL of QIAEX II matrix (Qiagen). The eluate was digested with NotI (reaction conditions: 4.5 μg of DNA, 50 units of NotI [New England Biolabs], 1× NEB3 Buffer in 150 μL volume, 22 h at 37°C), and the band centered at 520 bp was purified from a 12% polyacrylamide gel. The fragment was then concatenated by ligation (reaction conditions: 500 ng of DNA, 1× NEB T4 DNA ligase buffer, 200 units of T4 DNA ligase [New England Biolabs] in a 5-μL volume) and stored at 4°C. Prior to sonication, the ligation product was cleaned with QAIEX II and eluted in a 20-μL volume. After 20 min sonication, the sample was loaded on a 8% polyacrylamide gel, and the fraction from 100–300 bp was excised, purified, and blunted (reaction conditions: 1× NEB Blunting Buffer, 100 μM dNTPs, 1 μL of Blunting Enzyme Mix [New England Biolabs E1201S] in a 25-μL volume) for 30 min at 21°C. The product was purified by phenol/chloroform extraction and ethanol precipitation prior to A-tailing (reaction conditions: 5 units of Klenow Fragment [3′→5′ $exo^-$] [New England Biolabs], 1× reaction buffer, 200 μM dATP in a 50-μL volume, 30 min at 37°C). The product was purified by phenol/chloroform extraction and ethanol precipitation in preparation for ligation to Illumina TS adapters (www.illumina.com) (reaction conditions: 1× NEB T4 DNA ligase buffer, 1200 units of T4 DNA ligase [New England Biolabs], 1 μL of TS adapters in a 30-μL volume, 15 min at 21°C). The product was purified using a QiaQuick column (Qiagen) and eluted in a volume of 30 μL. Ten milliliters was then amplified by PCR using Illumina primers 1.1 and 2.2 (reaction conditions: 1 unit of enzyme, 2× *Pfx* amplification buffer, 1 mM $MgSO_4$, 0.3 μM oligonucleotides, and 0.3 mM each dNTP in a 25-μL volume. Two minute denaturation at 94°C was followed by 15 cycles of 30 sec at 94°C, 30 sec at 65°C, and 30 sec at 68°C, plus a final extension of 5 min at 68°C.). The PCR product was purified using a MinElute column (Qiagen) with a final volume of 13 μL.

### Illumina sequencing and analysis

We generated 18.8 million 36-nt reads and 21.7 million 50-nt single ends reads with the Illumina GAII analyzer, using sequencing chemistry version 2 (Illumina FC-204-2036) and cleavage reagent version 2 (Illumina #1005159). The combined set of 40,582,229 short reads were assembled using iSSAKE (with parameters: −m 15 −o 2 −r 0.7), as previously described (Warren et al. 2009). Briefly, reads aligning to the end of Ensembl *TRBV* gene predictions (Flicek et al. 2008) and having consecutively three or more unmatched bases in the adjacent $CDR3_\beta$ were used to seed iSSAKE de novo assemblies of $CDR3_\beta$. Only contigs with a depth of at least two reads at each position were retained for analysis. However, we did not set a requirement of double coverage for the seed sequences themselves. Given that a seed sequence could be long enough to span the complete $CDR3_\beta$, and there is no requirement for redundant coverage of seed sequences, 21,973 of the 33,664 clonotypes in our data set are represented by a single sequence read. This could, in principle, artifactually inflate the diversity of the repertoire, but in reality there is probably very little influence given that previous benchmarking of our assembly method using simulated, error-prone short sequence reads from 1 million computationally modeled $TCR_\beta$ sequences showed 93% sensitivity and 99.96% accuracy for $CDR3_\beta$ clonotypes present at >3 p.p.m. (Warren et al. 2009).

The iSSAKE contigs ($TCR_\beta$ reconstructions) were searched for the presence of 15 consecutive *TRBJ* segment bases. For any *TRBJ* segment, any 15-letter word from base position 1 to 25 characterizes uniquely that segment and allows the identification of the precise *TRBJ* segment boundary as well as the number of *TRBJ* bases deleted. *TRBV* segments, *TRBV* segment boundaries, and the exact number of deleted *TRBV* bases were inferred by tracing back the seed alignments that yielded the contigs and singlets under scrutiny. Sequence clonotypes were identified by extracting the contiguous bases spanning the last 15 bases of *TRBV* to the first recognizable 15 *TRBJ* segment bases, inclusively. During this automated process, we tracked clonotypes originating from seeds that aligned equally well to more than one *TRBV* segment, checked the sequence frame and peptide translation of the $TCR_\beta$ reconstructions, and extracted the $CDR3_\beta$, if applicable. The mined data was written to file and organized into a MySQL relational database for further analysis.

Fine-resolution analysis of N-diversity mechanisms at the V-D-J junction was made possible by searching for *TRBD1* (12 nt) and *TRBD2* (16 nt) bases between the *TRBV* and *TRBJ* boundaries using the longest to shortest *TRBD* word sizes. To favor accuracy over yield (some bases deletion/addition yield no recognizable *TRBD* bases), we chose to search until a minimum word of 8 nt for both *TRBD* segments. Unambiguous detection of *TRBD* bases allows precise identification of *TRBD* segment boundaries, the characterization of nontemplated bases at the V-D and D-J junction, and the frequency calculation of *TRBD* deleted bases for all clonotypes.

### Sanger sequencing and analysis

Purified fragment was inserted into the vector pCR-4 using the recommended conditions for Invitrogen's Zero Blunt TOPO PCR Cloning Kit for Sequencing and One Shot MAX Efficiency DH5α-T1R competent cells. M13FP and M13RP were used to prime Sanger Sequencing reactions. We generated paired reads from 384 clones.

Low-quality bases were trimmed and vector sequences screened using Cross_match (www.phrap.org). A total of 736 quality and vector-trimmed paired-end reads remained and were assembled using CAP3 (Huang and Madan 1999). Five hundred eighty-two reads (331 clones) contained the complete V(D)J sequence and collapsed into 220 unique contigs (including single-read contigs, or singlets). The resulting contigs and singlets were analyzed for the frequency of predicted $V_\beta$ and 14 $J_\beta$ gene segments. Briefly, we aligned contigs and singlets against two separate databases of Ensembl (Flicek et al. 2008) human gene predictions for 54 $V_\beta$ and 14 $J_\beta$ gene segments using WU-BLAST (default parameters with −b 3000 −v 3000). For each of the database sequences we tallied the $V_\beta$ and/or $J_\beta$ gene alignments having the highest sequence identity to the contig and singlet sequences. Sequence alignments were analyzed using custom scripts, noting both the exact position of each 3′-$V_\beta$ and/or 5′-$J_\beta$ segment onto the mRNA, and a report of $V_\beta$ frequency and $J_\beta$ frequency was generated. Mapping of both the $V_\beta$ and $J_\beta$ segment positions and ensuring that the translation frame was preserved and consistent with peptide predictions permitted the extraction of variable $CDR3_\beta$ bases between the two segments (not shown).

## Acknowledgments

This work was funded by Genome Canada and Genome British Columbia. R.A.H. is a Michael Smith Foundation for Health

Research scholar. We thank Martin Krzywinski for generating Figure 4 using the Circos software package (Krzywinski et al. 2009). The melanoma-specific T-cell clone DES M26 was kindly provided by Cassian Yee, Fred Hutchinson Cancer Research Centre, Seattle, WA.

## References

Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P. 1999. A direct estimate of the human alphabeta T cell receptor diversity. *Science* **286:** 958–961.

Bassing CH, Swat W, Alt FW. 2002. The mechanism and regulation of chromosomal V(D)J recombination. *Cell* (Suppl.) **109:** S45–S55.

Boria I, Cotella D, Dianzani I, Santoro C, Sblattero D. 2008. Primer sets for cloning the human repertoire of T cell receptor variable regions. *BMC Immunol* **9:** 50. doi: 10.1186/1471-2172-9-50.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: A sequence logo generator. *Genome Res* **14:** 1188–1190.

Davis MM, Bjorkman PJ. 1988. T-cell antigen receptor genes and T-cell recognition. *Nature* **334:** 395–402.

Douek DC, Betts MR, Brenchley JM, Hill BJ, Ambrozak DR, Ngai K, Karandikar NJ, Casazza JP, Koup RA. 2002. A novel approach to the analysis of specificity, clonality, and frequency of HIV-specific T cell responses reveals a potential mechanism for control of viral escape. *J Immunol* **168:** 3099–3104.

Even J, Lim A, Puisieux I, Ferradini L, Dietrich PY, Toubert A, Hercend T, Triebel F, Pannetier C, Kourilsky P. 1995. T-cell repertoires in healthy and diseased human tissues analysed by T-cell receptor beta-chain CDR3 size determination: Evidence for oligoclonal expansions in tumours and inflammatory diseases. *Res Immunol* **146:** 65–80.

Flicek P, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T, et al. 2008. Ensembl 2008. *Nucleic Acids Res* **36:** D707–D714.

Folch G, Lefranc, MP. 2000. The human T cell receptor beta variable (TRBV) genes. *Exp Clin Immunogenet* **17:** 42–54.

Fuschiotti P, Pasqual N, Hierle V, Borel E, London J, Marche PN, Jouvin-Marche E. 2007. Analysis of the TCR α-chain rearrangement profile in human T lymphocytes. *Mol Immunol* **44:** 3380–3388.

Gellert M. 1992. Molecular analysis of V(D)J recombination. *Annu Rev Genet* **26:** 425–446.

Gellert M. 2002. V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem* **71:** 101–132.

Gorski J, Yassai M, Zhu X, Kissela B, Keever C, Flomenberg N. 1994. Circulating T cell repertoire complexity in normal individuals and bone marrow recipients analyzed by CDR 3 size spectratyping: Correlation with immune status. *J Immunol* **152:** 5109–5119.

Hall MA, Lanchbury JS. 1995. Healthy human T-cell receptor β-chain repertoire quantitative analysis and evidence for Jβ-related effects on CDR3 structure and diversity. *Hum Immunol* **43:** 207–218.

Harty JT, Badovinac VP. 2008. Shaping and reshaping CD8 T cell memory. *Nat Rev Immunol* **8:** 107–119.

Holt RA, Jones SJ. 2008. The new paradigm of flow cell sequencing. *Genome Res* **18:** 839–846.

Huang X, Madan A. 1999. CAP3: A DNA sequence assembly program. *Genome Res* **9:** 868–877.

Jung D, Alt FW. 2004. Unraveling V(D)J recombination: Insights into gene regulation. *Cell* **116:** 299–311.

Krangel MS. 2003. Gene segment selection in V(D)J recombination: Accessibility and beyond. *Nat Immunol* **4:** 624–630.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: An information aesthetic for comparative genomics. *Genome Res* (this issue). doi: 10.1101/gr.092759.109.

Leslie A, Price DA, Mkhize P, Bishop K, Rathod A, Day C, Crawford H, Honeyborne I, Asher TE, Luzzi G, et al. 2006. Differential selection pressure exerted on HIV by CTL targeting identical epitopes but restricted by distinct HLA alleles from the same HLA supertype. *J Immunol* **177:** 4699–4708.

Li S, Wilkinson MF. 1998. Nonsense surveillance in lymphocytes? *Immunity* **8:** 135–141.

Manfras BJ, Terjung D, Boehm BO. 1999. Non-productive human TCR β chain genes represent V-D-J diversity before selection upon function: Insight into biased usage of TCRBD and TCRBJ genes and diversity of CDR3 region length. *Hum Immunol* **60:** 1090–1100.

Murphy KM, Travers P, Walport, M. 2007. *Janeway's immunobiology.* Garland Science, London, UK.

Nikolich-Zugich J, Slifka MK, Messaoudi I. 2004. The many important facets of T-cell repertoire diversity. *Nat Rev Immunol* **4:** 123–132.

Ozawa T, Tajiri K, Kishi H, Muraguchi A. 2008. Comprehensive analysis of the functional TCR repertoire at the single-cell level. *Biochem Biophys Res Commun* **367:** 820–825.

Pannetier C, Cochet M, Darche S, Casrouge A, Zoller M, Kourilsky P. 1993. The sizes of the CDR3 hypervariable regions of the murine T-cell receptor beta chains vary as a function of the recombined germ-line segments. *Proc Natl Acad Sci* **90:** 4319–4323.

Peters DG, Kassam AB, Yonas H, O'Hare EH, Ferrell RE, Brufsky AM. 1999. Comprehensive transcript analysis in small quantities of mRNA by SAGE-lite. *Nucleic Acids Res* **27:** e39.

Roldan EQ, Sottini A, Bettinardi A, Albertini A, Imberti L, Primi D. 1995. Different TCRBV genes generate biased patterns of VDJ diversity in human T cells. *Immunogenetics* **41:** 91–100.

Rosenberg WM, Moss PA, Bell JI. 1992. Variation in human T cell receptor V beta and J beta repertoire: Analysis using anchor polymerase chain reaction. *Eur J Immunol* **22:** 541–549.

Schneider TD, Stephens RM. 1990. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res* **18:** 6097–6100.

Shendure J, Ji H. 2008. Next-generation DNA sequencing. *Nat Biotechnol* **26:** 1135–1145.

Venturi V, Price DA, Douek DC, Davenport MP. 2008. The molecular basis for public T-cell responses? *Nat Rev Immunol* **8:** 231–238.

Warren RL, Nelson BH, Holt RA. 2009. Profiling model T cell metagenomes with short reads. *Bioinformatics* **25:** 458–464.

*Received February 18, 2009; accepted in revised form June 9, 2009.*

**Title:** Polyfunctional T-cell responses are disrupted by the ovarian cancer ascites environment and only partially restored by clinically relevant cytokines

**Authors and Affiliations:** Eric Tran[1,2], Julie S. Nielsen[2,4], Darin A. Wick[2], Alvin V. Ng[2], Nancy J. Nesslinger[2], Elissa McMurtrie[3], John R. Webb[1,2], and Brad H. Nelson[1,2,4]

[1]Department of Microbiology and Biochemistry, University of Victoria, Victoria BC, Canada, [2]Trev and Joyce Deeley Research Centre, British Columbia Cancer Agency, Victoria BC, Canada, [3]BC Cancer Agency, Victoria, BC, Canada, and [4]Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada.

**Corresponding Author:** Brad H. Nelson, Trev and Joyce Deeley Research Centre, BC Cancer Agency, 2410 Lee Ave., Victoria, BC, V8R 6V5, Canada. Phone: 1-250-519-5705; Fax: 1-250-519-2004; E-mail: bnelson@bccancer.bc.ca

**Abstract**

**Background:** Host T-cell responses are associated with favorable outcomes in epithelial ovarian cancer (EOC), but it remains unclear how best to promote these responses in patients. Toward this goal, we evaluated a panel of clinically relevant cytokines for the ability to enhance multiple T-cell effector functions (polyfunctionality) in the native tumor environment. **Methodology/Principle Findings:** Experiments were performed with resident CD8+ and CD4+ T cells in bulk ascites cell preparations from high-grade serous EOC patients. T cells were stimulated with α-CD3 in the presence of 100% autologous ascites fluid with or without exogenous IL-2, IL-12, IL-18 or IL-21, alone or in combination. T-cell proliferation (Ki-67) and function (IFN-γ, TNF-α, IL-2, CCL4, and CD107a expression) were assessed by multi-parameter flow cytometry. We found that ascites fluid had variable effects on CD8+ and CD4+ T-cell proliferation, but inhibited T-cell function in most patient samples, with CD107a, IFN-γ, and CCL4 showing the greatest inhibition. T-cell proliferation was enhanced by exogenous IL-2, but other T-cell functions were largely unaffected by single cytokines. The combination of IL-2 with cytokines engaging complementary signaling pathways, in particular IL-12 and IL-18, enhanced expression of IFN-γ, TNF-α, and CCL4 in all patient samples by promoting polyfunctional T-cell responses. However, no combination of cytokines enhanced expression of CD107a or IL-2. **Conclusions/Significance:** The EOC ascites environment disrupts multiple T-cell functions, and exogenous cytokines engaging diverse signaling pathways only partially reverse these effects. Our results may explain the limited efficacy of cytokine therapies for EOC to date. Full restoration of T-cell

function will require activation of signaling pathways beyond those engaged by IL-2, IL-12, IL-18 and IL-21.

**Key Words:** T-cell therapy, immunotherapy, cytokines, polyfunctional, ovarian cancer

## Introduction

Numerous studies in the past decade have shown that the immune system influences clinical outcomes in patients with high-grade serous epithelial ovarian cancer (hereafter referred to as EOC). Specifically, the presence of tumor-infiltrating CD8+ T cells is associated with prolonged disease-free and overall survival [1,2]. Moreover, elevated numbers of CD56+ T cells in the ascites compartment are correlated with increased platinum sensitivity [3]. Together, these studies suggest that enhancing the endogenous T-cell response to EOC may be of clinical benefit to patients. However, the ovarian tumor environment contains many immunosuppressive cell types and factors that can oppose anti-tumor T-cell responses. Accordingly, elevated numbers of tumor-associated regulatory T cells (Tregs) and macrophages are associated with poor survival in EOC [1,2]. Many soluble immunosuppressive factors are also found in ascites, including IL-10, TGF-β, VEGF, B7-H1/PD-L1, B7-H4, SDF-1, EBAG9/RCAS1, Fas ligand, and soluble IL-2 receptor [1,2].

The heterogeneity of immunosuppressive mechanisms in EOC presents a formidable challenge for immunotherapy, as it suggests that multiple immunosuppressive mechanisms might need to be reversed to restore T-cell function. For example, Treg depletion is being evaluated as a means to enhance immunity against several human cancers [4], including EOC [5], but at best would only reverse one mechanism of immunosuppression, leaving other immunological barriers intact. Rather than attempt to disable these barriers one by one, a more pragmatic clinical approach would be to deliver factors that can broadly override immunosuppressive barriers in the tumor environment.

Thus, identifying the signals or conditions that promote T-cell responses in the EOC environment may lead to more effective immunotherapeutic strategies.

To this end, we previously developed a mouse model of EOC that allows functional assessment of CD8+ T cell responses in the ovarian tumor environment [6]. Specifically, we engineered a murine EOC tumor line to express a CD8+ T cell epitope from the model antigen ovalbumin. Mice bearing advanced, widely disseminated tumors with extensive ascites were treated by adoptive transfer of CD8+ T cells specific for the ovalbumin epitope. Remarkably, adoptively transferred CD8+ T cells underwent vigorous proliferation not only in lymph nodes and peripheral blood, but also in the ascites and tumor compartments. Indeed, at the peak of the response, the transferred T cells constituted up to 40% of CD8+ T cells in peripheral blood, and up to 96% of CD8+ T cells in ascites. This profound proliferative response was followed by rapid and near-complete tumor regression in the majority of animals, demonstrating that the T cells were functionally active. Importantly, the proliferation and anti-tumor activity of the transferred CD8+ T cells was entirely dependent on IL-2/IL-15 signaling, as demonstrated by genetic disruption of the IL-2 receptor alpha (CD25) or beta (CD122) subunits [6]. Thus, even in the setting of advanced ovarian tumors, CD8+ T cells mounted potent anti-tumor responses provided that IL-2/IL-15 signaling pathways were intact.

The foregoing results raised the issue of whether cytokines such as IL-2 can similarly over-ride the immunosuppressive effects of ascites in human EOC. Previous *in vitro* studies with human EOC samples have shown that IL-2 can partially restore lymphokine-activated killer (LAK) cell cytotoxicity in the presence of 50% ascites fluid,

while the combination of IL-2 with TCR stimulation [7] or IL-12 [8,9] fully restored LAK cytotoxicity. However, use of 50% ascites fluid alone does not fully recapitulate the ovarian tumor environment, which also contains immunosuppressive cell types such as Tregs and MDSCs [1,2]. Furthermore, LAKs are predominately comprised of NK cells, which may be affected differently by ascites compared to T cells. These past studies also measured cytotoxicity by bulk LAK cell preparations and hence did not elucidate the effects of cytokines on different T cell subsets. Finally, *in vitro* cytotoxicity is only one aspect of T-cell function and does not always correlate with effective T-cell responses *in vivo* [10].

There is growing appreciation of the importance of polyfunctional T-cells, i.e., T cells that can simultaneously perform multiple functions, in effective immune responses [11]. Specifically, polyfunctional T-cell responses are associated with protective immunity after vaccination against smallpox (vaccinia virus) [12], yellow fever [13], tuberculosis [14,15], and leishmaniasis [16]. Elevated numbers of polyfunctional T cells are also correlated with favorable outcomes in a variety of disease settings, including HIV/AIDS [17,18,19,20,21,22,23,24,25,26,27], hepatitis C [28] and lymphocytic choriomeningitis [29]. In the setting of cancer, several studies have found enhanced numbers of tumor-specific polyfunctional T cells in patients responding favorably to various forms of immunotherapy, including adoptive T-cell therapy, α-CTLA antibody therapy, and a multi-peptide + DNA vaccine [30,31,32,33,34]. However, little is known about the effects of the EOC environment on polyfunctional T-cell responses.

In the present study, we used multiparameter flow cytometry to assess the effects of the human EOC ascites environment on multiple T-cell functions, including expression

of IFN-γ, TNF-α, IL-2, CCL4, and surface localization of CD107a (a marker of degranulation). Furthermore, we assessed the ability of IL-2 and three other clinically relevant cytokines (IL-12, IL-18 and IL-21) to restore T-cell function in the ascites tumor environment. We found that (a) ascites inhibited some T-cell functions (CD107a, CCL4, and IFN-γ) more strongly than others (proliferation, TNF-α, and IL-2); (b) exogenous IL-2 generally enhanced T-cell proliferation but had little effect on other T-cell functions; and (c) the combination of IL-2 with cytokines engaging complementary signaling pathways (in particular IL-12 and IL-18) enhanced proliferation and expression of IFN-γ, CCL4 and TNF-α but failed to enhance CD107a or IL-2. Thus, T-cell function in the human EOC environment can be partially restored by exogenous cytokines. However, full restoration of T-cell function will require activation of signaling pathways beyond those engaged by IL-2, IL-12, IL-18 and IL-21.

## Results

### EOC ascites disrupts multiple T-cell functions

Primary ascites specimens were collected from patients with high-grade serous EOC (Table 1). To model the native ascites tumor environment as best as possible *in vitro*, we analyzed bulk ascites cell pellets, which in addition to CD8+ and CD4+ T cells, contained variable numbers of tumor cells, Tregs, B cells, NK cells, and CD14+ cells (Table 2). Moreover, cells were cultured in 100% autologous ascites fluid to include any soluble factors, such as TGF-β (Table 2), that may impact T-cell function. Because of the paucity of well-defined T-cell antigens in ovarian cancer, we stimulated cultures with α-CD3 antibody, which should stimulate both effector and regulatory subsets. Thus, each patient's T cells were analyzed in the presence of the full complement of naturally occurring tumor cells, immunosuppressive cells, and soluble factors.

We recently demonstrated in a murine model of EOC that the ascites environment can be highly permissive to T-cell proliferation [6]. To see if this was also true in human EOC samples, we tested the effect of ascites fluid on T-cell proliferation. As seen in Fig. 1A, ascites fluid had highly variable effects on CD8+ T-cell proliferation (as measured by Ki-67 expression), ranging from strong inhibition (1/5 patients) to enhancement of proliferation (2/5 patients). In general, similar trends were seen for CD4+ T cells (Fig. 1A, and data not shown). Thus, as suggested by our murine studies [6], the human EOC ascites environment is not universally suppressive and in some cases can enhance T-cell proliferation.

We next determined whether other T-cell functions were similarly affected by ovarian ascites. Flow cytometric analysis was used to assess five commonly used markers of T-cell function: IFN-γ, TNF-α, IL-2, CCL4, and CD107a [11,12,17,18,19]. The first four markers are cytokines involved in T-cell proliferation or effector function, whereas CD107a is expressed on the surface of T cells undergoing cellular degranulation and thus serves as a marker of cytotoxicity [11]. In all patient samples tested, ascites fluid caused a dramatic inhibition of CD107a expression by CD8+ T cells, suggesting that the ascites environment generally inhibits T-cell degranulation (Fig. 1A). CCL4 and IFN-γ expression were also inhibited by ascites in the majority of patient samples (Fig. 1A). T cells expressing IL-2 and TNF-α were relatively rare, and were minimally affected by ascites fluid in most cases (Fig. 1A). Similar results were seen with CD4+ T cells (data not shown). Intriguingly, the effects of ascites on T-cell proliferation and other T-cell functions were largely uncoupled. For example, with patient samples IROC008 and IROC036, ascites fluid enhanced both CD8+ and CD4+ T-cell proliferation but inhibited expression of CD107a, CCL4, and IFN-γ (Fig. 1A, and data not shown). Thus, ascites fluid has widely variable effects on T-cell functions between patient samples, in accord with the heterogenous nature of EOC.

Although a systematic evaluation of the myriad immunosuppressive factors in EOC [1,2] was beyond the scope of this study, we did investigate the possible influence of Tregs and TGF-β. The percentage of Tregs ranged from 2.5% to 6.8% of total cells but showed no correlation with the degree of inhibition of T-cell proliferation or other functions (Table 2; Pearson correlation, $p > 0.05$ for all comparisons). Likewise, TGF-β levels did not correlate with the extent of T-cell inhibition (Table 2; Pearson correlation p

> 0.05 for all comparisons). These findings underscore the complex immunobiology of the ovarian tumor environment in that the degree of immunosuppression did not correlate with either one of these common immunosuppressive factors.

**Exogenous cytokines only partially restore T-cell function in the EOC ascites environment**

In our mouse model of EOC, IL-2/IL-15 signaling was critical for CD8+ T-cell proliferation and function in ascites [6]. However, the preceding experiments revealed that the vast majority of T cells in EOC ascites did not express detectable levels of IL-2. This led us to speculate that the addition of IL-2 to cultures might restore T-cell proliferation and function. To test this, bulk ascites cells were stimulated with α-CD3 in the presence of 100% ascites fluid with or without exogenous IL-2, and T-cell proliferation and function were measured 48h later. In 4/5 patient samples, IL-2 fully reversed the effects of ascites on T-cell proliferation and indeed induced responses greater than those seen in complete media (Fig. 1B). By contrast, IL-2 had little effect on the expression of CD107a, CCL4, IFN-γ, or IL-2 for most patient samples, although it caused a weak but statistically significant enhancement of TNF-α expression (Fig. 1B). Thus, IL-2 alone enhanced T-cell proliferation but, in general, had negligible effects on other functions.

We next asked if the engagement of other signaling pathways might better enhance T-cell function. Whereas IL-2 activates primarily the STAT5 pathway, the related cytokine IL-12 activates the STAT4 pathway, which has been shown to enhance T

cell function in various physiological settings [36]. When added to EOC cultures, IL-12 alone had only minor effects on the proliferation or function of CD8+ T cells (Fig. 1B) in most patient samples. We next evaluated IL-21, which predominantly activates the STAT1 and STAT3 pathways [37]. Similar to IL-12, IL-21 had only modest effects on T-cell proliferation or function (Fig. 1B). In general, similar trends were seen with CD4+ T cells (data not shown).

Reasoning that a broader degree of functional enhancement might be achieved by simultaneously activating multiple STAT pathways, we assessed two combinations of cytokines: a) IL-2 + IL-12, and b) IL-2 + IL-12 + IL-21. Importantly, the latter combination would theoretically activate all STAT pathways relevant to CD8+ T-cell responses (i.e., STAT1, STAT3, STAT4 and STAT5). In general, both cytokine combinations induced T-cell proliferation similar to that seen with IL-2 (compare Fig. 2 with Fig. 1B). Notably however, CD8+ T cells from IROC028, which failed to proliferate in response to any single cytokine (Fig. 1B), proliferated in response to both cytokine combinations (Fig. 2). Moreover, both cytokine combinations were modestly more effective than single cytokines at enhancing IFN-$\gamma$ and TNF-$\alpha$ production (Fig. 2). Despite this, neither cytokine combination enhanced expression of CD107a or IL-2 (Fig. 2). Similar results were seen with CD4+ T cells, although the effects of cytokines were generally weaker (data not shown). In summary, the addition of cytokine combinations engaging a wide range of STAT signaling pathways failed to fully restore T-cell function in the EOC ascites environment.

Finally, we assessed whether activation of non- STAT signaling pathways could further enhance T-cell function. To this end, we evaluated the cytokine IL-18, which

belongs to the IL-1 cytokine family, activates the MyD88/NFκb pathway, and can directly enhance CD8+ T-cell responses in a variety of settings [38,39,40]. IL-18 alone had weak effects on T-cell proliferation and function (Fig. 3A). However, when combined with IL-2 + IL-12, IL-18 significantly enhanced expression of IFN-γ and, to a lesser extent, TNF-α (Fig. 3A). Moreover, the combination of IL-2 + IL-12 + IL-18 significantly increased the amount of IFN-γ produced on a per-cell basis, as evidenced by increased mean fluorescence intensity (MFI) (Fig. 3B). Nonetheless, the combination of IL-2 + IL-12 + IL-18 was unable to significantly enhance expression of CD107a or IL-2 (Fig. 3A). Similar results were seen for CD4+ T cells, although IL-18 did not increase the MFI of IFN-γ (data not shown).

**Exogenous cytokines induce new polyfunctional T-cell profiles**

In theory, the enhanced activity of cytokine combinations compared to single cytokines could reflect either (a) different cytokines stimulating distinct subpopulations of T cells, or (b) cytokines acting synergistically to induce single T cells to perform multiple functions (i.e., induction of polyfunctional T cells). To investigate this issue, we used Boolean gate analysis to study the different functional permutations displayed by CD8+ and CD4+ T cells under different stimulation conditions. As seen in the top panel of Fig. 4, when bulk ascites cells were stimulated in complete media, only five predominant functional permutations were seen among CD8+ T cells: (a) non-functional (i.e., zero function) T cells (data not shown); (b) mono-functional T cells expressing CCL4 or CD107a (permutations 2 and 5), (c) bi-functional T cells expressing both

CD107a and CCL4 (permutation 13), and (d) tri-functional T cells expressing CD107a, CCL4, and IFN-γ (permutation 24). In the presence of 100% ascites fluid, the number of non-functional T cells increased, whereas the other functional permutations were reduced with the exception of mono-functional T cells expressing CCL4 (Fig. 4, second panel). In general, addition of single cytokines failed to significantly change the functional permutations seen with ascites alone, although IL-2 had weak effects on several permutations (Supplementary Fig. 1). In contrast, the addition of cytokine combinations (either IL-2 + IL-12 + IL-21, or IL-2 + IL-12 + IL-18) markedly decreased the number of zero- and monofunctional T cells and generated three new polyfunctional permutations: a) CCL4 and IFN-γ (permutation 10); b) CCL4, IFN-γ, and TNF-α (permutation 17); and c) CCL4, IFN-γ, TNF-α, and CD107a (permutation 28) (Fig. 4, third and fourth panel). Similar trends were seen for CD4+ T cells, although the effects were generally weaker (data not shown). Thus, these combinations of cytokines appear to act synergistically to induce polyfunctional responses by individual T cells as opposed to stimulating multiple subpopulations of mono-functional T cells.

## Discussion

We show here that the ascites environment in human EOC has variable effects on CD8+ and CD4+ T-cell proliferation and indeed can even enhance proliferation in some cases. By contrast, ascites generally inhibits expression of IFN-γ, TNF-α, CCL4, and CD107a by T cells, resulting in a preponderance of zero- or monofunctional T cells. Although IL-2 promoted T-cell proliferation, it was largely ineffective at enhancing other T-cell functions, as were other single agent cytokines (IL-12, IL-18 and IL-21). In contrast, combinations of cytokines that activate complementary signaling pathways (in particular IL-2 + IL-12 + IL-18) were able to enhance expression of IFN-γ and, to a lesser extent, TNF-α, and CCL4. Nevertheless, expression of IL-2 and CD107a were unaffected by any combination of cytokines. Thus, exogenous cytokines can partly restore T-cell function in the human EOC environment; however, alternative strategies will be required to fully restore T-cell function.

One limitation of this study is the relatively small sample size, which was necessary given the number of cytokines and functional read-outs involved. Despite the small sample size, there was a notable degree of similarity between the overall functional profiles of all five patient samples both at baseline and after cytokine stimulation. This suggests that the diverse immunosuppressive mechanisms present in ovarian cancer might ultimately converge toward a common T-cell functional state governed by a conserved set of regulatory signals.

The inability of the clinical cytokines and cytokine combinations tested here to enhance T-cell degranulation and IL-2 production in the EOC ascites environment raises

important considerations for the cytokine therapy of ovarian cancer. Given that granule exocytosis is a major mechanism by which T cells kill virus-infected and transformed cells [41], the inability of T cells to degranulate is expected to significantly impair anti-tumor responses. Intriguingly, whereas we found that cytokines were ineffective at reversing ascites-mediated inhibition of degranulation, previous studies have shown that the combination of IL-2 with TCR stimulation [7] or IL-12 [8,9] could fully restore LAK cytotoxicity in the presence of 50% ascites fluid. This discrepancy could reflect differences in cell culture conditions, cell type analyzed, and/or the method used to assess cytotoxic status. In the prior studies with LAK cells [7,8,9], perhaps the observed cytotoxic activity was mediated by the Fas/FasL pathway rather than cytotoxic granule release. The inability of T cells to produce IL-2 is also expected to impair the anti-tumor response. In our mouse model of ovarian cancer [6], tumor regression was entirely dependent on IL-2/IL-15 signaling. Likewise, Gattinoni and colleagues demonstrated that effector CD8+ T cells capable of potent *in vitro* anti-tumor cytotoxicity and IFN-γ production, but not IL-2 production, had limited anti-tumor activity *in vivo*, whereas T cells that had low *in vitro* cytotoxicity, but high IL-2 production, mediated superior anti-tumor responses [10]. As CD8+ T cells generally lose the ability to produce IL-2 upon differentiation, the failure of cytokines to enhance IL-2 production suggests that most tumor-associated T cells may be terminally differentiated effectors [11]. If so, future efforts should be directed toward isolating and expanding less differentiated subsets of T cells with greater pluripotency.

Even though the combination of IL-2 + IL-12 + IL-18 did not rescue expression of CD107a and IL-2, it potently enhanced the amount of IFN-γ produced by individual

CD8+ T cells (Fig. 3B). IFN-γ plays a central role in anti-tumor immunity by inhibiting tumor proliferation and angiogenesis, and upregulating tumor antigen presentation. Endogenous IFN-γ protects against tumor development, and in a number of tumor models, is critical for anti-tumor immunity [42]. Moreover, IFN-γ and IFN-γ receptor expression, as well as several genes downstream of IFN-γ, are associated with increased survival in human EOC [1]. How might IL-2, IL-12 and IL-18 cooperatively enhance IFN-γ production? First, these cytokines reciprocally upregulate expression of their respective receptors, resulting in enhanced sensitivity of T cells to all three cytokines [43,44,45,46]. Second, these cytokines can mediate synergistic and complementary signaling events [36,47,48]. For example, IL-2 and IL-12 synergistically activate the p38 MAPK pathway to enhance IFN-γ expression [49]. Likewise, IL-12 and IL-18 activate the transcription factors STAT4 and AP-1, respectively, which can synergistically enhance IFN-γ promoter activity [50]. Thus, the coordinated activation of both STAT and non-STAT signaling pathways appears important for maximal expression of IFN-γ and other effector cytokines.

The combination of IL-2 + IL-12 + IL-18 also induced significant changes in polyfunctional T-cell permutations. Although the clinical significance of these polyfunctional profiles has yet to be determined for EOC, insights can be gained from other disease settings. Two of the four functional permutations seen at baseline (specifically, monofunctional CD107a, and bi-functional CD107a and CCL4; Fig. 4) are also seen elevated in CD8+ T cells from HIV progressors compared to non-progressors, suggesting that these permutations may represent functionally exhausted T cells [17,18]. Addition of IL-2 + IL-12 + IL-18 decreased the number of CD8+ T cells exhibiting these

"exhausted" functional permutations. Moreover, it caused the emergence of three new polyfunctional permutations: a) CCL4 and IFN-γ; b) CCL4, IFN-γ, and TNF-α; and c) CCL4, IFN-γ, TNF-α, and CD107a (Fig. 4). Notably, these three permutations were shown to be induced in CD8+ T cells by a protective vaccine against smallpox [12]. Furthermore, tetra-functional (CCL4, IFN-γ, TNF-α, and CD107a) CD8+ T cells are found at significantly higher frequencies in HIV long-term non-progressors compared to progressors [17,18]. Finally, a higher proportion of tumor-reactive T cells co-expressing CCL4, IFN-γ, and TNF-α were detected in melanoma patients with favorable clinical responses to α-CTLA-4 treatment [32]. These prior reports suggest that the polyfunctional permutations induced by IL-2 + IL-12 + IL-18 may have therapeutic benefit in EOC, although this suggestion needs to be validated *in vivo*.

Our findings shed new light on previous clinical trials in EOC involving IL-2 and IL-12. Intraperitoneal administration of IL-2 resulted in an approximately 25 % overall response rate in two studies [51,52], while IL-2 in combination with retinoic acid as a maintenance therapy demonstrated a relatively modest, but statistically significant, prolongation of progression-free and overall survival in a phase II trial [53,54]. By contrast, IL-12 has shown little therapeutic efficacy in EOC in two clinical trials [55,56]. The limited efficacy of IL-2 and IL-12 as monotherapies is consistent with our results demonstrating the modest functional activity of these cytokines as single agents. Our results would also predict that IL-18 and IL-21 may have limited clinical efficacy when used as monotherapies. Indeed, although IL-18 and IL-21 have not been evaluated in EOC patients, these cytokines as single agents showed limited clinical benefit in the setting of metastatic melanoma [57,58,59,60,61,62]. When combined, IL-2, IL-12, and

IL-18 showed the most promise in the *in vitro* experiments reported here, but unfortunately these cytokines would likely have significant side effects if given in combination to patients [45,63,64]. While it may be possible to reduce toxicity by optimizing dose and route of delivery (e.g., intraperitoneal administration) [45,65], one is still left with the fact that this combination does not fully restore T-cell function, as manifested by impaired expression of CD107a and IL-2. Thus, our findings indicate that alternative therapeutic strategies involving other signaling pathways will be required to unleash the full potential of host T-cell responses against EOC.

## Materials and Methods

### Study subjects and specimens.

**Ethics Statement.** Newly diagnosed patients with high-grade serous EOC gave written informed consent under protocols approved by the Research Ethics Board of the BC Cancer Agency and the University of British Columbia.

Tumor tissue and ascites were obtained during primary surgery prior to any other treatment. Ascites was centrifuged at 300 g, and supernatants (ascites fluid) were stored at -80°C. Ascites cell (AC) pellets containing large quantities of red blood cells were treated with ACK lysis buffer. AC pellets were cryopreserved in liquid nitrogen. Upon thawing, ascites cells were rested in complete RPMI (RPMI 1640 with 10% FBS, 25mM HEPES, 1mM sodium pyruvate, 2mM L-glutamine, and 50μm β-mercaptoethanol) for 4h at 37°C prior to experiments.

**Antibodies and cytokines.** Flow cytometry was performed with the following fluorochrome-conjugated antibodies (BD Biosciences): CD3 (FITC, PECy5), CD4 (FITC, PE), CD8 (PECy5, APC-H7), CD14 (PE), CD19 (PE), CD25 (PECy5), CD56 (PE), CD107a (PECy5), Ki-67 (FITC), CCL4 (Mip-1β, PE), IL-2 (APC), TNF-α (PECy7), and IFN-γ (PE, AlexaFluor 700). Foxp3 (PE) was from eBioscience. Recombinant human IL-12, IL-21 (both Peprotech), and IL-18 (R&D Systems) were used

418 at 100ng/ml. IL-2 (Proleukin) was used at 100U/ml. TGF-β (latent and active) in ascites
419 fluid was quantified by ELISA (eBioscience).
420
421 **Proliferation assays.** Ascites cells (AC) were seeded in triplicate in 96-well flat bottom
422 plates at 1 x $10^5$ cells per well. AC were resuspended in media or 100% ascites fluid and
423 were left unstimulated or stimulated with plate bound α-CD3ε antibody (clone OKT3,
424 eBioscience) previously coated at 2.5μg/ml for 2h at 37°C, in the presence or absence of
425 cytokines. T-cell proliferation was measured by detection of Ki-67 using flow cytometry.
426 Cells were washed with FACS buffer (1% FBS in PBS), permeabilized with ice-cold
427 100% methanol and incubated at -20°C for at least 1h. Cells were then washed with
428 FACS buffer and triple-stained with pre-titered antibodies to Ki-67, CD4, and CD8 for at
429 least 30min at room temperature in the dark. Cells were washed and analyzed with a
430 FACSCalibur flow cytometer (Becton Dickinson). Data was analyzed using FlowJo
431 software (Tree Star Inc.).
432
433 **Assessment of T-cell functional markers.** Bulk ascites cells were washed with serum-
434 free RPMI, resuspended at 5 x $10^5$ cells/ml in complete RPMI or autologous ascites fluid,
435 and plated in 48-well plates at 3 x $10^5$ cells/well with or without plate bound α-CD3ε.
436 Cytokines and α-CD107a antibody were added to appropriate wells, and cells were
437 incubated for 48h at 37°C. As per standard protocol, α-CD107a antibody was added
438 during stimulation because CD107a, which is associated with the membranes of cytotoxic
439 granules, transiently localizes to the surface of T cells undergoing cellular degranulation
440 [35]. Thus, cell-surface expression of CD107a serves as a surrogate marker of

cytotoxicity. As a positive control, PMA plus ionomycin (both from Sigma) were used at 40ng/ml and 1.5μg/ml, respectively. GolgiStop (BD Bioscience, 1:1500 dilution) and brefeldinA (Sigma, 5μg/ml) were added for the final 6h of culture. Cells were harvested, labeled with α-CD4 and CD8 antibodies, fixed and permeabilized with Cytofix/Cytoperm buffer (BD Bioscience) according to manufacturer's instructions, washed with Perm/Wash buffer, and stored in FACS buffer overnight at 4°C. Cells were washed again with Perm/Wash buffer, and intracellular staining was performed using antibodies against CCL4, IL-2, TNF-α, and IFN-γ. Cells were fixed with 2% formaldehyde and stored in FACS buffer overnight at 4°C. Samples were analyzed with a BD Bioscience FACSVantage DIVA modified with the Octagon array. SSC Area vs. SSC W were used to gate out doublets. Data was analyzed with FlowJo software (TreeStar) and exported to PESTLE v1.6.1 (Mario Roederer, NIH) for further data analysis.

**Acknowledgements**

We thank the patients who generously donated specimens; Dr. Peter Watson, Jodi LeBlanc, Rebecca Barnes, Sindy Babinszky, and Kristy Dillon for specimen and clinical data collection; Dr. Julian Lum, David McPhee, Dr. Cecile LePage, Rob Sahota, and Katy Milne for insightful comments; and Dr. Mario Roederer for kindly providing the PESTLE v1.6.1 software.

## References


1. Nelson BH (2008) The impact of T-cell immunity on ovarian cancer outcomes. Immunol Rev 222: 101-116.
2. Chu CS, Kim SH, June CH, Coukos G (2008) Immunotherapy opportunities in ovarian cancer. Expert Rev Anticancer Ther 8: 243-257.
3. Bamias A, Tsiatas ML, Kafantari E, Liakou C, Rodolakis A, et al. (2007) Significant differences of lymphocytes isolated from ascites of patients with ovarian cancer compared to blood and tumor lymphocytes. Association of CD3(+)CD56(+) cells with platinum resistance. Gynecol Oncol.
4. Zou W (2006) Regulatory T cells, tumour immunity and immunotherapy. Nat Rev Immunol 6: 295-307.
5. Barnett B, Kryczek I, Cheng P, Zou W, Curiel TJ (2005) Regulatory T cells in ovarian cancer: biology and therapeutic potential. Am J Reprod Immunol 54: 369-377.
6. Yang T, Wall EM, Milne K, Theiss P, Watson P, et al. (2007) CD8+ T Cells Induce Complete Regression of Advanced Ovarian Cancers by an Interleukin (IL)-2/IL-15 Dependent Mechanism. Clin Cancer Res 13: 7172-7180.
7. Hirte HW, Clark DA, O'Connell G, Rusthoven J, Mazurka J (1992) Reversal of suppression of lymphokine-activated killer cells by transforming growth factor-beta in ovarian carcinoma ascitic fluid requires interleukin-2 combined with anti-CD3 antibody. Cell Immunol 142: 207-216.
8. Barton DP, Blanchard DK, Duan C, Roberts WS, Cavanagh D, et al. (1995) Interleukin-12 synergizes with interleukin-2 to generate lymphokine-activated killer activity in peripheral blood mononuclear cells cultured in ovarian cancer ascitic fluid. J Soc Gynecol Investig 2: 762-771.
9. DeCesare SL, Michelini-Norris B, Blanchard DK, Barton DP, Cavanagh D, et al. (1995) Interleukin-12-mediated tumoricidal activity of patient lymphocytes in an autologous in vitro ovarian cancer assay system. Gynecol Oncol 57: 86-95.
10. Gattinoni L, Klebanoff CA, Palmer DC, Wrzesinski C, Kerstann K, et al. (2005) Acquisition of full effector function in vitro paradoxically impairs the in vivo antitumor efficacy of adoptively transferred CD8+ T cells. J Clin Invest 115: 1616-1626.
11. Seder RA, Darrah PA, Roederer M (2008) T-cell quality in memory and protection: implications for vaccine design. Nat Rev Immunol 8: 247-258.
12. Precopio ML, Betts MR, Parrino J, Price DA, Gostick E, et al. (2007) Immunization with vaccinia virus induces polyfunctional and phenotypically distinctive CD8(+) T cell responses. J Exp Med 204: 1405-1416.
13. Gaucher D, Therrien R, Kettaf N, Angermann BR, Boucher G, et al. (2008) Yellow fever vaccine induces integrated multilineage and polyfunctional immune responses. J Exp Med 205: 3119-3131.
14. Beveridge NE, Price DA, Casazza JP, Pathan AA, Sander CR, et al. (2007) Immunisation with BCG and recombinant MVA85A induces long-lasting, polyfunctional Mycobacterium tuberculosis-specific CD4+ memory T lymphocyte populations. Eur J Immunol 37: 3089-3100.
15. Elvang T, Christensen JP, Billeskov R, Thi Kim Thanh Hoang T, Holst P, et al. (2009) CD4 and CD8 T cell responses to the M. tuberculosis Ag85B-TB10.4

promoted by adjuvanted subunit, adenovector or heterologous prime boost vaccination. PLoS ONE 4: e5139.
16. Darrah PA, Patel DT, De Luca PM, Lindsay RW, Davey DF, et al. (2007) Multifunctional TH1 cells define a correlate of vaccine-mediated protection against Leishmania major. Nat Med 13: 843-850.
17. Betts MR, Nason MC, West SM, De Rosa SC, Migueles SA, et al. (2006) HIV nonprogressors preferentially maintain highly functional HIV-specific CD8+ T cells. Blood 107: 4781-4789.
18. Duvall MG, Precopio ML, Ambrozak DA, Jaye A, McMichael AJ, et al. (2008) Polyfunctional T cell responses are a hallmark of HIV-2 infection. Eur J Immunol 38: 350-363.
19. Almeida JR, Price DA, Papagno L, Arkoub ZA, Sauce D, et al. (2007) Superior control of HIV-1 replication by CD8+ T cells is reflected by their avidity, polyfunctionality, and clonal turnover. J Exp Med 204: 2473-2485.
20. Almeida JR, Sauce D, Price DA, Papagno L, Shin SY, et al. (2009) Antigen sensitivity is a major determinant of CD8+ T-cell polyfunctionality and HIV suppressive activity. Blood 113: 6351-6360.
21. Ferre AL, Hunt PW, Critchfield JW, Young DH, Morris MM, et al. (2009) Mucosal immune responses to HIV-1 in elite controllers: a potential correlate of immune control. Blood 113: 3978-3989.
22. Rehr M, Cahenzli J, Haas A, Price DA, Gostick E, et al. (2008) Emergence of polyfunctional CD8+ T cells after prolonged suppression of human immunodeficiency virus replication by antiretroviral therapy. J Virol 82: 3391-3404.
23. Kannanganat S, Kapogiannis BG, Ibegbu C, Chennareddi L, Goepfert P, et al. (2007) Human immunodeficiency virus type 1 controllers but not noncontrollers maintain CD4 T cells coexpressing three cytokines. J Virol 81: 12071-12076.
24. Bansal A, Jackson B, West K, Wang S, Lu S, et al. (2008) Multifunctional T-cell characteristics induced by a polyvalent DNA prime/protein boost human immunodeficiency virus type 1 vaccine regimen given to healthy adults are dependent on the route and dose of administration. J Virol 82: 6458-6469.
25. Valor L, Navarro J, Carbone J, Rodriguez-Sainz C, Gil J, et al. (2008) Immunization with an HIV-1 immunogen induces CD4+ and CD8+ HIV-1-specific polyfunctional responses in patients with chronic HIV-1 infection receiving antiretroviral therapy. Vaccine 26: 2738-2745.
26. Winstone N, Guimaraes-Walker A, Roberts J, Brown D, Loach V, et al. (2009) Increased detection of proliferating, polyfunctional, HIV-1-specific T cells in DNA-modified vaccinia virus Ankara-vaccinated human volunteers by cultured IFN-gamma ELISPOT assay. Eur J Immunol 39: 975-985.
27. Lichterfeld M, Yu XG, Waring MT, Mui SK, Johnston MN, et al. (2004) HIV-1-specific cytotoxicity is preferentially mediated by a subset of CD8(+) T cells producing both interferon-gamma and tumor necrosis factor-alpha. Blood 104: 487-494.
28. Badr G, Bedard N, Abdel-Hakeem MS, Trautmann L, Willems B, et al. (2008) Early interferon therapy for hepatitis C virus infection rescues polyfunctional, long-lived CD8+ memory T cells. J Virol 82: 10017-10031.

552 29. Yi JS, Du M, Zajac AJ (2009) A Vital Role for Interleukin-21 in the Control of a
553 Chronic Viral Infection. Science 324: 1572-1576.
554 30. Imai N, Ikeda H, Tawara I, Shiku H (2009) Tumor progression inhibits the induction
555 of multifunctionality in adoptively transferred tumor-specific CD8+ T cells. Eur J
556 Immunol 39: 241-253.
557 31. Imai N, Ikeda H, Tawara I, Wang L, Wang L, et al. (2009) Glucocorticoid-induced
558 tumor necrosis factor receptor stimulation enhances the multifunctionality of
559 adoptively transferred tumor antigen-specific CD8 T cells with tumor regression.
560 Cancer Sci 100: 1317-1325.
561 32. Yuan J, Gnjatic S, Li H, Powel S, Gallardo HF, et al. (2008) CTLA-4 blockade
562 enhances polyfunctional NY-ESO-1 specific T cell responses in metastatic
563 melanoma patients with clinical benefit. Proc Natl Acad Sci U S A 105: 20410-
564 20415.
565 33. Carpenito C, Milone MC, Hassan R, Simonet JC, Lakhal M, et al. (2009) Control of
566 large, established tumor xenografts with genetically retargeted human T cells
567 containing CD28 and CD137 domains. Proc Natl Acad Sci U S A 106: 3360-
568 3365.
569 34. Perales MA, Yuan J, Powel S, Gallardo HF, Rasalan TS, et al. (2008) Phase I/II study
570 of GM-CSF DNA as an adjuvant for a multipeptide cancer vaccine in patients
571 with advanced melanoma. Mol Ther 16: 2022-2029.
572 35. Betts MR, Brenchley JM, Price DA, De Rosa SC, Douek DC, et al. (2003) Sensitive
573 and viable identification of antigen-specific CD8+ T cells by a flow cytometric
574 assay for degranulation. J Immunol Methods 281: 65-78.
575 36. Watford WT, Hissong BD, Bream JH, Kanno Y, Muul L, et al. (2004) Signaling by
576 IL-12 and IL-23 and the immunoregulatory roles of STAT4. Immunol Rev 202:
577 139-156.
578 37. Zeng R, Spolski R, Casas E, Zhu W, Levy DE, et al. (2007) The molecular basis of
579 IL-21-mediated proliferation. Blood 109: 4135-4142.
580 38. Iwai Y, Hemmi H, Mizenina O, Kuroda S, Suda K, et al. (2008) An IFN-gamma-IL-
581 18 signaling loop accelerates memory CD8+ T cell proliferation. PLoS ONE 3:
582 e2404.
583 39. Carroll RG, Carpenito C, Shan X, Danet-Desnoyers G, Liu R, et al. (2008) Distinct
584 effects of IL-18 on the engraftment and function of human effector CD8 T cells
585 and regulatory T cells. PLoS One 3: e3289.
586 40. Okamoto I, Kohno K, Tanimoto T, Ikegami H, Kurimoto M (1999) Development of
587 CD8+ effector T cells is differentially regulated by IL-18 and IL-12. J Immunol
588 162: 3202-3211.
589 41. Cullen SP, Brunet M, Martin SJ Granzymes in cancer and immunity. Cell Death
590 Differ 17: 616-623.
591 42. Ikeda H, Old LJ, Schreiber RD (2002) The roles of IFN gamma in protection against
592 tumor development and cancer immunoediting. Cytokine Growth Factor Rev 13:
593 95-109.
594 43. Ahn HJ, Maruo S, Tomura M, Mu J, Hamaoka T, et al. (1997) A mechanism
595 underlying synergy between IL-12 and IFN-gamma-inducing factor in enhanced
596 production of IFN-gamma. Journal of Immunology 159: 2125-2131.

44. Smeltz RB, Chen J, Hu-Li J, Shevach EM (2001) Regulation of interleukin (IL)-18 receptor alpha chain expression on CD4(+) T cells during T helper (Th)1/Th2 differentiation. Critical downregulatory role of IL-4. J Exp Med 194: 143-153.
45. Weiss JM, Subleski JJ, Wigginton JM, Wiltrout RH (2007) Immunotherapy of cancer by IL-12-based cytokine combinations. Expert Opin Biol Ther 7: 1705-1721.
46. Xu D, Chan WL, Leung BP, Hunter D, Schulz K, et al. (1998) Selective expression and functions of interleukin 18 receptor on T helper (Th) type 1 but not Th2 cells. Journal of Experimental Medicine 188: 1485-1492.
47. Gracie JA, Robertson SE, McInnes IB (2003) Interleukin-18. J Leukoc Biol 73: 213-224.
48. Nelson BH, Willerford DM (1998) Biology of the interleukin-2 receptor. Adv Immunol 70: 1-81.
49. Gollob JA, Schnipper CP, Murphy EA, Ritz J, Frank DA (1999) The functional synergy between IL-12 and IL-2 involves p38 mitogen-activated protein kinase and is associated with the augmentation of STAT serine phosphorylation. Journal of Immunology 162: 4472-4481.
50. Nakahira M, Ahn HJ, Park WR, Gao P, Tomura M, et al. (2002) Synergy of IL-12 and IL-18 for IFN-gamma gene expression: IL-12-induced STAT4 contributes to IFN-gamma promoter activation by up-regulating the binding activity of IL-18-induced activator protein 1. J Immunol 168: 1146-1153.
51. Vlad AM, Budiu RA, Lenzner DE, Wang Y, Thaller JA, et al. (2010) A phase II trial of intraperitoneal interleukin-2 in patients with platinum-resistant or platinum-refractory ovarian cancer. Cancer Immunol Immunother 59: 293-301.
52. Edwards RP, Gooding W, Lembersky BC, Colonello K, Hammond R, et al. (1997) Comparison of toxicity and survival following intraperitoneal recombinant interleukin-2 for persistent ovarian cancer after platinum: twenty-four-hour versus 7-day infusion. J Clin Oncol 15: 3399-3407.
53. Recchia F, Di Orio F, Candeloro G, Guerriero G, Piazze J, et al. Maintenance immunotherapy in recurrent ovarian cancer: long term follow-up of a phase II study. Gynecol Oncol 116: 202-207.
54. Recchia F, Saggio G, Cesta A, Candeloro G, Nuzzo A, et al. (2005) Interleukin-2 and 13-cis retinoic acid as maintenance therapy in advanced ovarian cancer. Int J Oncol 27: 1039-1046.
55. Lenzi R, Edwards R, June C, Seiden MV, Garcia ME, et al. (2007) Phase II study of intraperitoneal recombinant interleukin-12 (rhIL-12) in patients with peritoneal carcinomatosis (residual disease < 1 cm) associated with ovarian cancer or primary peritoneal carcinoma. J Transl Med 5: 66.
56. Hurteau JA, Blessing JA, DeCesare SL, Creasman WT (2001) Evaluation of recombinant human interleukin-12 in patients with recurrent or refractory ovarian cancer: a gynecologic oncology group study. Gynecol Oncol 82: 7-10.
57. Tarhini AA, Millward M, Mainwaring P, Kefford R, Logan T, et al. (2009) A phase 2, randomized study of SB-485232, rhIL-18, in patients with previously untreated metastatic melanoma. Cancer 115: 859-868.
58. Robertson MJ, Kirkwood JM, Logan TF, Koch KM, Kathman S, et al. (2008) A dose-escalation study of recombinant human interleukin-18 using two different

schedules of administration in patients with cancer. Clin Cancer Res 14: 3462-3469.

59. Robertson MJ, Mier JW, Logan T, Atkins M, Koon H, et al. (2006) Clinical and biological effects of recombinant human interleukin-18 administered by intravenous infusion to patients with advanced cancer. Clin Cancer Res 12: 4265-4273.

60. Davis ID, Brady B, Kefford RF, Millward M, Cebon J, et al. (2009) Clinical and biological efficacy of recombinant human interleukin-21 in patients with stage IV malignant melanoma without prior treatment: a phase IIa trial. Clin Cancer Res 15: 2123-2129.

61. Davis ID, Skrumsager BK, Cebon J, Nicholaou T, Barlow JW, et al. (2007) An open-label, two-arm, phase I trial of recombinant human interleukin-21 in patients with metastatic melanoma. Clin Cancer Res 13: 3630-3636.

62. Thompson JA, Curti BD, Redman BG, Bhatia S, Weber JS, et al. (2008) Phase I study of recombinant interleukin-21 in patients with metastatic melanoma and renal cell carcinoma. J Clin Oncol 26: 2034-2039.

63. Osaki T, Peron JM, Cai Q, Okamura H, Robbins PD, et al. (1998) IFN-gamma-inducing factor/IL-18 administration mediates IFN-gamma- and IL-12-independent antitumor effects. Journal of Immunology 160: 1742-1749.

64. Carson WE, Dierksheide JE, Jabbour S, Anghelina M, Bouchard P, et al. (2000) Coadministration of interleukin-18 and interleukin-12 induces a fatal inflammatory response in mice: critical role of natural killer cell interferon-gamma production and STAT-mediated signal transduction. Blood 96: 1465-1473.

65. Osaki T, Hashimoto W, Gambotto A, Okamura H, Robbins PD, et al. (1999) Potent antitumor effects mediated by local expression of the mature form of the interferon-gamma inducing factor, interleukin-18 (IL-18). Gene therapy 6: 808-815.

680 **Figure Legends**

681

682 **Fig. 1. Impact of ascites fluid and IL-2, IL-12, and IL-21 on CD8+ T-cell**
683 **proliferation and function.** Bulk ascites cells were stimulated with plate bound α-CD3
684 in media or 100% autologous ascites fluid for 48h. (**A**) Proliferation and function of
685 CD8+ T cells in the presence of media or autologous ascites fluid (top panel) was
686 assessed by measuring expression of Ki-67, CD107a, CCL4, IFN-γ, TNF-α, and IL-2 by
687 flow cytometry. (**B**) Effect of IL-2 (second panel), IL-12 (third panel), and IL-21 (fourth
688 panel) on CD8+ T-cell proliferation and various functions in the presence of autologous
689 ascites fluid. Data has been normalized to α-CD3 stimulated cells in media (i.e., media
690 values were subtracted from each stimulation condition). *There was a significant effect
691 of the cytokine for enhancing the indicated function compared to cells stimulated in
692 media (Wilcoxon matched pairs t test, $p < 0.05$).

693

694 **Fig. 2. Effects of cytokine combinations on CD8+ T-cell proliferation and function.**
695 Bulk ascites cells were stimulated with plate bound α-CD3 in media or 100% autologous
696 ascites fluid for 48h. Proliferation and function of CD8+ T cells in the presence of media,
697 autologous ascites fluid (top panel), or ascites fluid in the presence of IL-2 + IL-12
698 (middle panel), or IL-2 + IL-12 + IL-21 (bottom panel) was assessed by measuring
699 expression of Ki-67, CD107a, CCL4, IFN-γ, TNF-α, and IL-2 by flow cytometry. Data
700 has been normalized to α-CD3 stimulated cells in media (i.e., media values were
701 subtracted from each stimulation condition). *There was a significant effect of the

cytokine combination for enhancing the indicated function compared to cells stimulated in media (Wilcoxon matched pairs t test, $p < 0.05$).

**Fig. 3. Effects of IL-18 alone, or in combination with IL-2 and IL-12, on CD8+ T-cell proliferation and function.** Bulk ascites cells were stimulated with plate bound α-CD3 in media or 100% autologous ascites fluid for 48h. Proliferation and function of CD8+ T cells in the presence of media or autologous ascites fluid (top panel) was assessed by measuring expression of Ki-67, CD107a, CCL4, IFN-γ, TNF-α, and IL-2 by flow cytometry. **(A)** Effect of IL-18 (middle panel) and IL-2 + IL-12 + IL-18 (bottom panel) on CD8+ T-cell proliferation and various functions in the presence of ascites fluid. Data has been normalized to α-CD3 stimulated cells in media (i.e., media values were subtracted from each stimulation condition). *There was a significant effect of IL-2 + IL-12 + IL-18 for enhancing the indicated function compared to cells stimulated in media (Wilcoxon matched pairs t test, $p < 0.05$). **(B)** Mean fluorescence intensity (MFI) of IFN-γ positive CD8+ T cells was determined by intracellular IFN-γ staining. *The effect of IL-2 + IL-12 + IL-18 was significantly greater than the other two cytokine combinations (Wilcoxon matched pairs t test, $p < 0.05$).

**Fig. 4. Effects of cytokine combinations on polyfunctional CD8+ T-cell responses.** Bulk ascites cells were stimulated with plate bound α-CD3 in media or 100% autologous ascites fluid for 48h in the presence or absence of the indicated cytokine or cytokine combination. Boolean gate analysis was performed to quantify the number of T cells expressing each of 31 possible functional permutations. Shown are the results for CD8+

T cells stimulated in media, ascites fluid, or ascites fluid supplemented with the cytokine combinations IL-2 + IL-12 + IL-21 or IL-2 + IL-12 + IL-18. The frequency of T cells expressing a given permutation is expressed as a percentage of total CD8+ T cells. *For the indicated permutation, the effect of the cytokine combination was significantly greater than that seen with media (Wilcoxon matched pairs t test, $p < 0.05$).

**Tables**

**Table 1. Patient clinical characteristics**

| Patient ID | Age at diagnosis | Pathologic diagnosis | Grade | FIGO Staging | TNM Staging |
|---|---|---|---|---|---|
| IROC008 | 70 | Papillary serous carcinoma | 3/3 | 4 | pT3c, pN1 |
| IROC028 | 61 | Papillary serous carcinoma | 3/3 | 3C | pT3c, NX, MX |
| IROC034 | 64 | Papillary serous carcinoma | 3/3 | N/A | T3c |
| IROC036 | 60 | Papillary serous carcinoma | 3/3 | N/A | T3c, NX, MX |
| IROC038 | 40 | Papillary serous carcinoma | 3/3 | 3B | pT3b |

Median Age: 61; Mean Age: 59; Age Range: 40-70

N/A, not assessed

**Table 2. Cellular composition and TGF-β levels in the ascites compartment of EOC patients (data is expressed as a percentage of total live cells unless otherwise stated)**

| Parameter | IROC Patient Sample | | | | |
|---|---|---|---|---|---|
| | **008** | **028[A]** | **034** | **036** | **038** |
| Lymphocytes[B] | 28.5 | 74.6 | 78.0 | 20.6 | 40.2 |
| CD3+ | 19.8 | 59.0 | 58.0 | 11.9 | 21.9 |
| CD3+CD4+ | 6.5 | 41.1 | 26.8 | 4.5 | 6.2 |
| CD3+CD8+ | 12.4 | 16.9 | 28.1 | 6.6 | 13.0 |
| CD3+CD56+ | 3.8 | 1.6 | 10.1 | 1.0 | 3.2 |
| CD56+ | 2.7 | 12.0 | 9.7 | 3.0 | 4.6 |
| Tregs[C] (CD4+CD25+FoxP3+) | 5.2 | 6.8 | 2.9 | 3.3 | 2.5 |
| CD25+FoxP3+ in CD4+ | 12.4 | 12.1 | 12.1 | 6.8 | 10.7 |
| B cells[C] (CD19+) | 8.1 | 1.4 | 6.5 | 4.5 | 5.0 |
| CD14+ | 55.3 | 11.2 | 1.6 | 50.5 | 30.9 |
| TGF-β (pg/ml) | 6.3 | ND[D] | 33.3 | ND | ND |

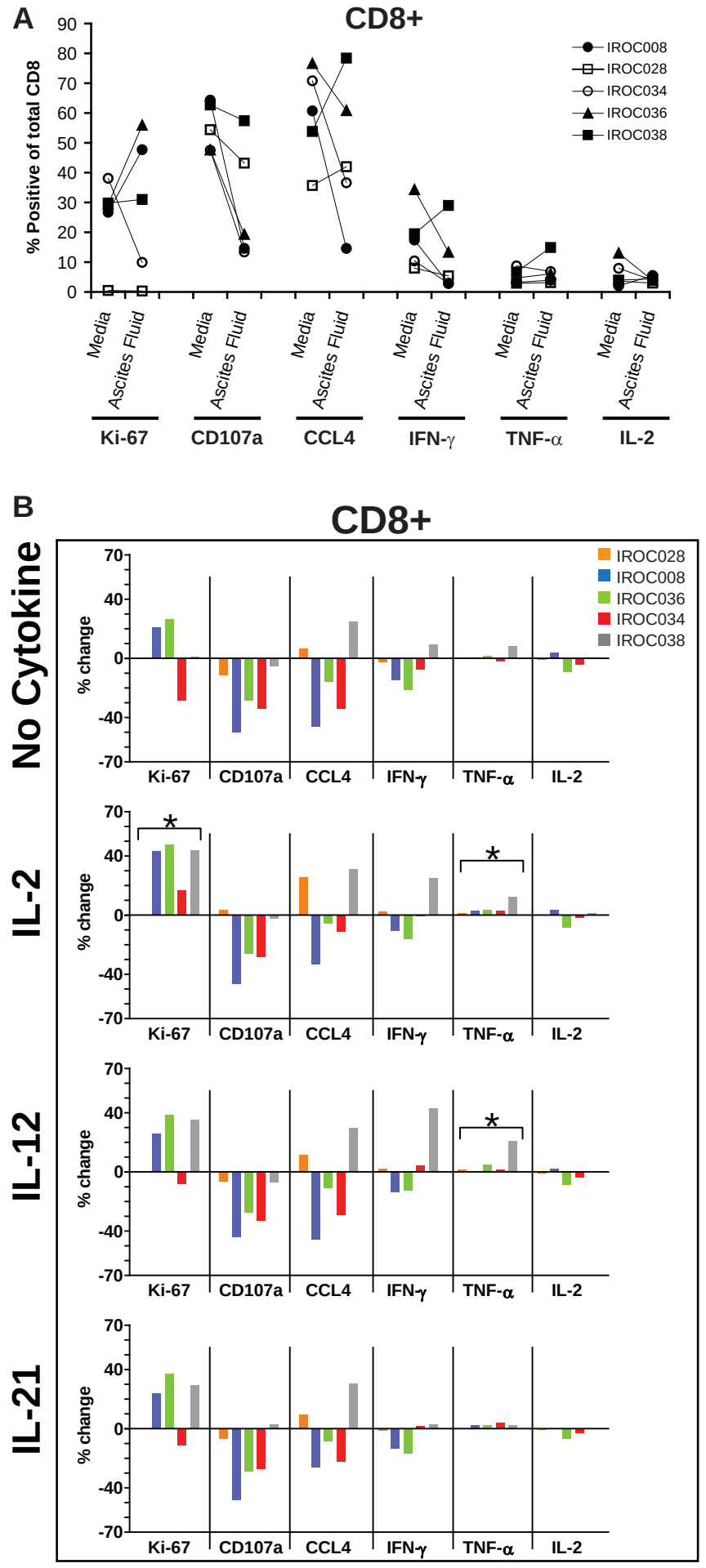
[A] Sample contained many large tumor rafts not detectable by FACS. Therefore, values are inflated since visual inspection and IHC staining shows this sample is comprised mainly of tumor cells.

[B] As determined by side and forward scatter.

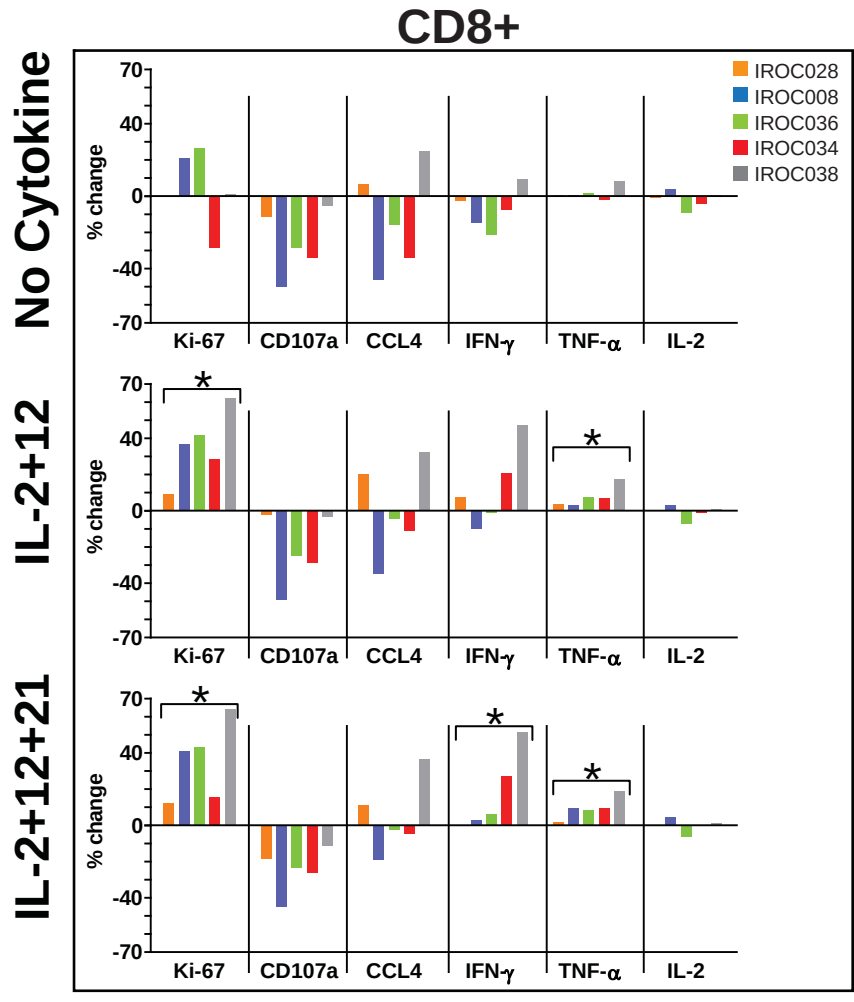
[C] Tregs and B cells expressed as a percentage of total live lymphocytes

[D] ND, not detectable

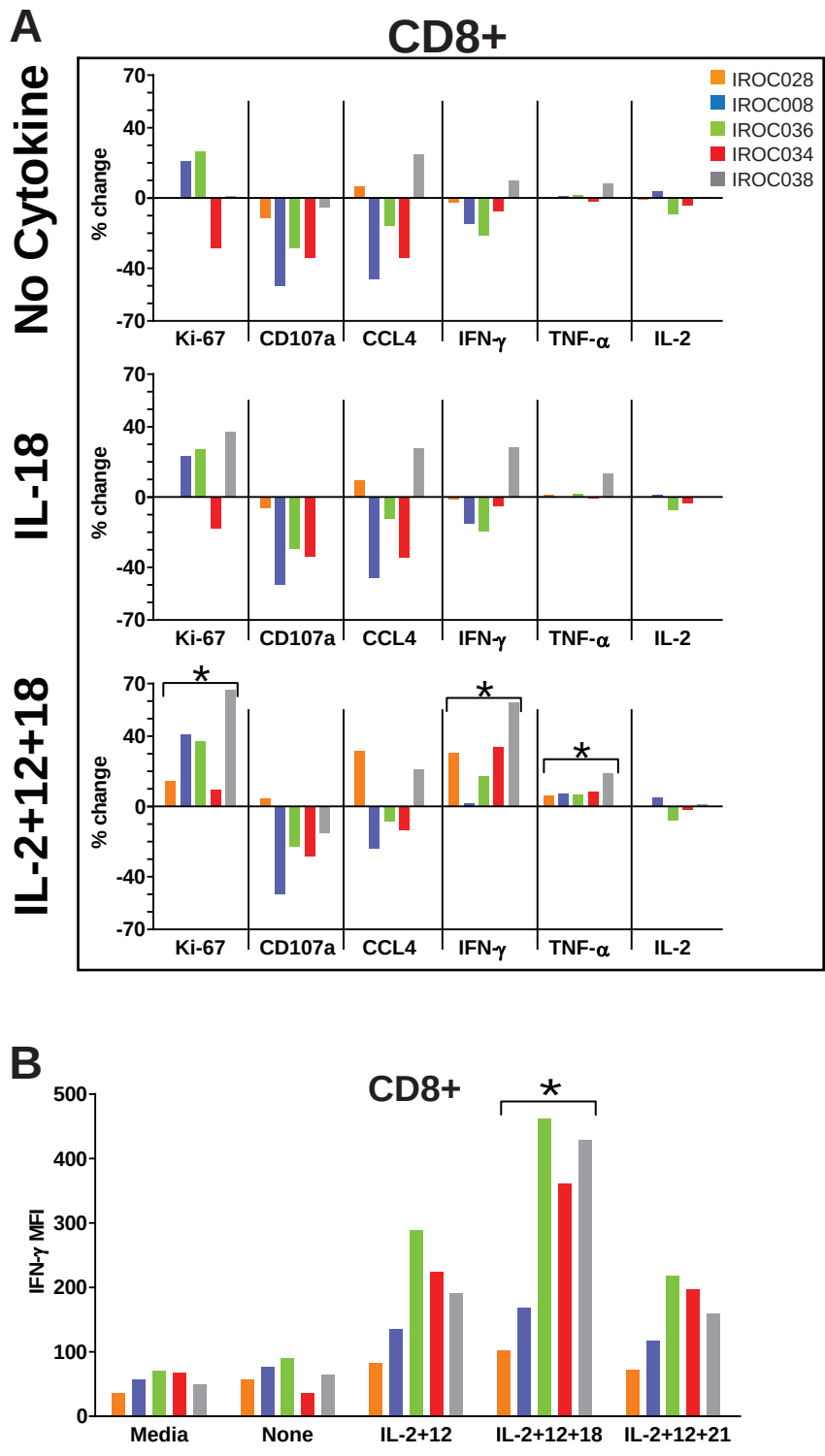
Tran et al. Fig. 1



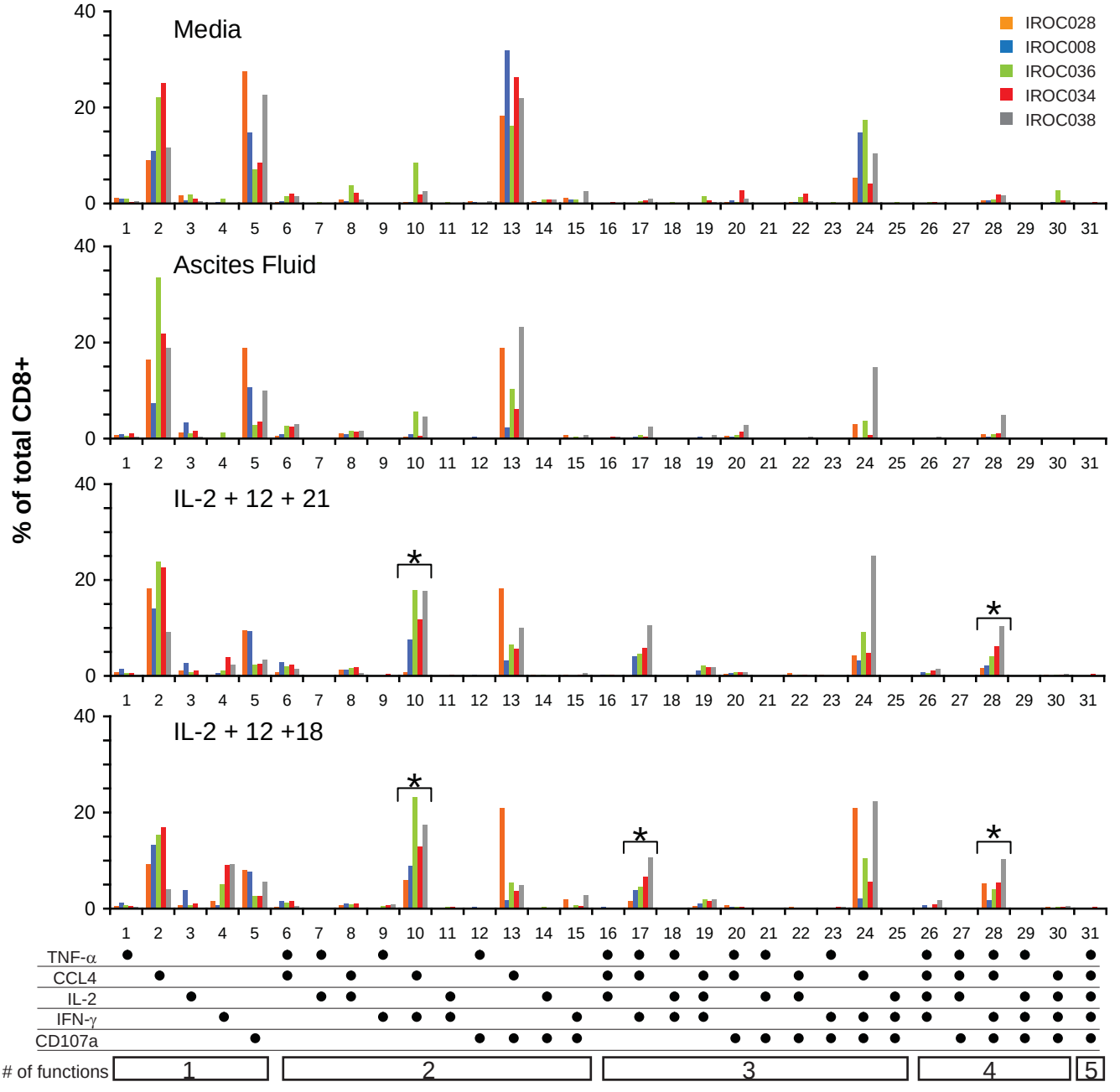
Tran et al. Fig. 2



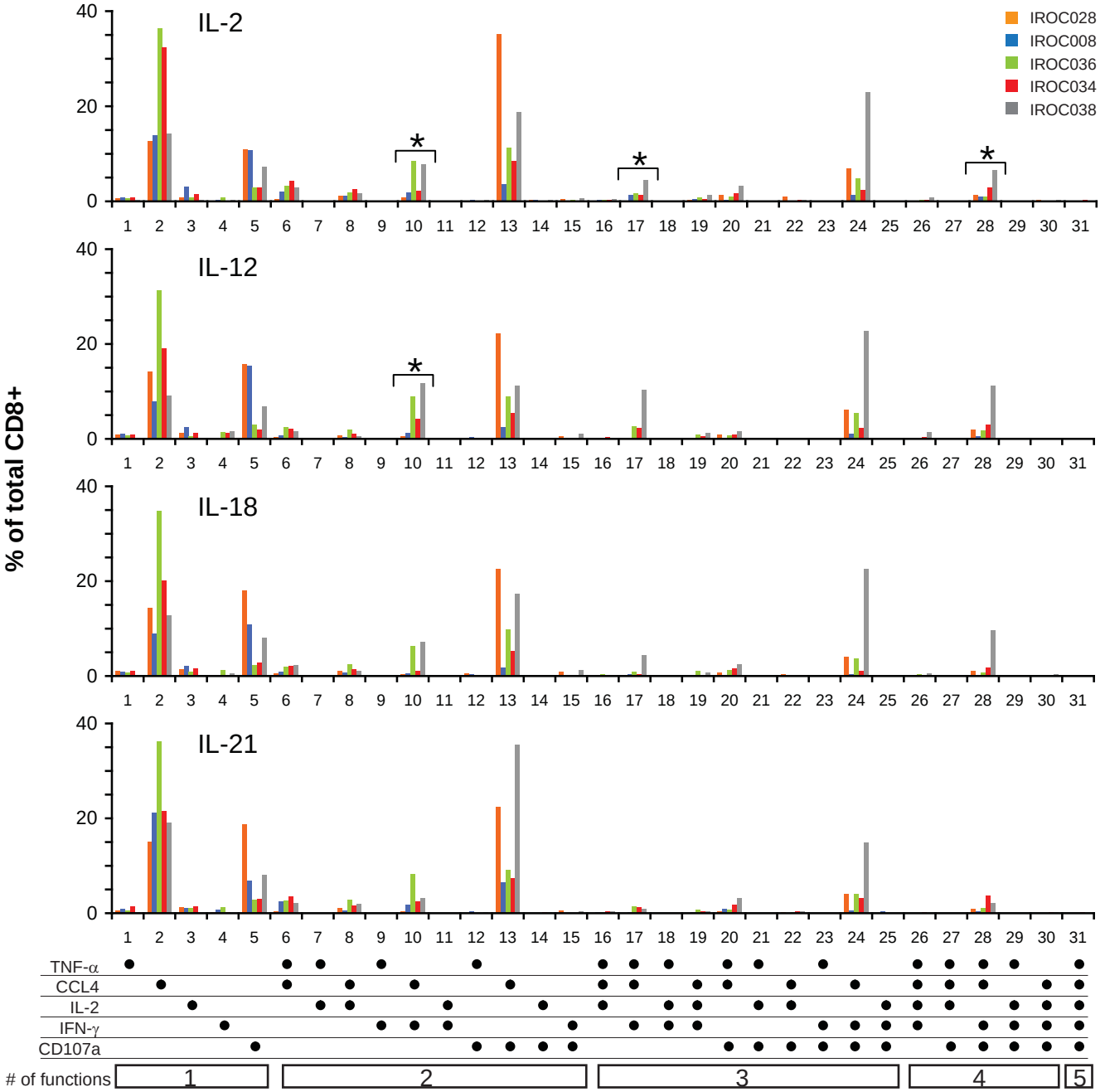
Tran et al. Fig. 3



Tran et al. Fig. 4



Tran et al. Supp Fig. 1



Gynecologic Oncology 115 (2009) 179–180

Contents lists available at ScienceDirect

Gynecologic Oncology

journal homepage: www.elsevier.com/locate/ygyno

Editorial

# IDO and outcomes in ovarian cancer

Ovarian cancer affects over 190,000 women worldwide each year (International Agency for Research on Cancer). While over 80% of patients are highly responsive to frontline treatment (cytoreductive surgery followed by taxane-and platinum-based chemotherapy), a large majority experience disease recurrence within 2–5 years and ultimately succumb to their disease. Despite these unfortunate statistics, 20–30% of ovarian cancer patients survive 5 years or more after diagnosis. Favorable prognostic factors include early stage, non-serous histology, low grade, good performance status, and optimal surgical debulking.

In the past few years, it has become apparent that the host immune system too has a strong influence on survival from ovarian cancer. Specifically, the presence of CD8+T cells in tumor epithelium has been associated with prolonged disease-free and overall survival in numerous studies (reviewed in [1]). Other features of CD8+T cell responses are also associated with favorable prognosis in ovarian cancer, including intratumoral levels of interferon-γ and its receptor, IL-18, TNF-α, IRF-1, MHC class I molecules and antigen processing machinery, and the cytolytic granule component TIA-1 [1]. Concordant results have been reported in a wide variety of other human cancers, leading to the general view that host T cell responses can profoundly influence clinical outcomes in human cancer.

In this issue of *Gynecological Oncology*, Inaba and colleagues [2] investigate another facet of the immune response to ovarian cancer. Using a cohort of 60 ovarian cancer cases representing a range of histological subtypes, stage and grade, they assessed expression of the enzyme indoleamine-2,3-dioxygenase (IDO) in ovarian tumors. They report that IDO expression is associated with high grade, significantly fewer intraepithelial CD8+T cell infiltrates, and decreased overall and progression-free survival. These findings are reminiscent of prior work by this group in endometrial cancer [3,4], and by others in serous ovarian cancer [5,6], colorectal cancer [7], and hepatocellular carcinoma [8], all indicating an inverse association between IDO expression and clinical outcomes.

The association between IDO expression and reduced T cell infiltrates fits with work by this group and others showing that IDO can have immunosuppressive effects [9]. Activated lymphocytes, by releasing IFN-γ, can induce IDO expression in a variety of tissues. In turn, IDO can inhibit T cell proliferation and function by several mechanisms. IDO degrades tryptophan, yielding breakdown products called kynurenines. Thus, expression of IDO can deplete tryptophan locally, leaving T cells starved for this amino acid. In addition, kynurenines can directly cause T cell apoptosis. Finally, regulatory T cells can induce expression of IDO on dendritic cells, which in turn can inhibit the activation of naive T cells in tumor draining lymph nodes. The relative importance of these different mechanisms to immune suppression is unclear, and likely depends on physiological context [9].

The influence of IDO on physiological immune responses also appears to depend on context. Several studies have demonstrated a clear immune suppressive role. Munn et al. [10] showed that IDO expression at the feto-maternal interface was crucial to prevent rejection of allogeneic fetuses in mice. In murine tumor models, forced expression of IDO protects tumors from T cell-mediated rejection. And in the allogeneic transplant setting, IDO-deficient mice fail to control lethal CD8+T cell responses. Despite these findings, IDO−/− mice have intact central and peripheral tolerance and do not develop autoimmunity. Moreover, in a study of human kidney allografts, IDO expression was seen in rejected organs only [11], indicating that IDO expression is not invariably associated with immune suppression.

In the present paper, Inaba and colleagues provide thought-provoking data about the role of IDO in the immune response to ovarian cancer. Based on their findings in human ovarian tumors, they transfected IDO into the ovarian cancer line SKOV3 and assessed the *in vitro* and *in vivo* consequences on cell behavior. When evaluated *in vitro*, IDO-expressing cells showed normal morphology, proliferation, migration, invasive activity, and sensitivity to the chemotherapeutic agent paclitaxel. However, when an IDO-expressing clone was injected inraperitoneally into nude mice, there was a marked increase in the volume and extent of dissemination of IDO-expressing tumors compared to the parental tumor line. This effect could be suppressed by daily administration of the IDO inhibitor 1-methyl-tryptophan (1-MT). While administration of 1-MT alone did not lead to increased survival of mice, 1-MT potentiated the therapeutic effects of paclitaxel, as reported previously by this group in a model of endometrial cancer [12].

While the therapeutic effects of 1-MT and paclitaxel speak for themselves, the immunological basis of these observations is unclear. Using a spontaneous mammary tumor model, others have shown that 1-MT potentiates the effects of paclitaxel by a T cell-dependent mechanism [13]. However, the present study used nude mice, which lack T cells. Nonetheless, the authors note that nude mice contain NK cells, which could be the target of immune suppression by IDO and rescue by 1-MT. Indeed, this group has presented evidence that IDO can suppress NK cell responses in nude mice bearing endometrial tumors [12].

Given the enhanced therapeutic effects of 1-MT in combination with paclitaxel, should IDO inhibitors be considered for the treatment of ovarian cancer? The strong association between tumor-infiltrating T cells and survival suggests that any intervention that enhances T cell immunity is likely to be clinically beneficial, provided side effects are low. However, several issues need to be resolved in the case of IDO inhibitors. First, there is controversy over the relative efficacy of the D

0090-8258/$ – see front matter 
doi:10.1016/j.ygyno.2009.09.017

and L isoforms in different models, an issue that needs clarification prior to human trials [9,14]. Second, more information is needed about the off-target effects of IDO inhibitors, such as the inhibition of tryptophan transporters, which can even affect IDO-negative cells. Third, another form of IDO was recently discovered (IDO2), and more study is required into the role of IDO2 in immune suppression and cancer progression [15]. Finally, the results by Inada and colleagues suggest there is more to be learned about the immunological effects of IDO inhibitors, given the therapeutic effects they observed in T cell-deficient hosts. In addition to the NK cell hypothesis they propose, perhaps IDO can promote tumor growth and dissemination by non-immunological mechanisms as well. For example, by depleting tryptophan, IDO could potentially induce an altered metabolic state in tumor cells that changes their growth and dissemination properties. Thus, the present paper highlights the importance of IDO in influencing ovarian cancer outcomes, and underscores the need to better understand the underlying mechanisms to allow successful translation of these findings to the clinic.

## References


[1] Nelson BH. The impact of T-cell immunity on ovarian cancer outcomes. Immunol Rev 2008 Apr;222:101–16.

[2] Inaba T, Ino K, Kajiyama H, Yamamoto E, Shibata K, Nawa A, et al. Role of the immunosuppressive enzyme indoleamine 2,3-dioxygenase in the progression of ovarian carcinoma. Gynecol Oncol 2009 Aug 7.

[3] Ino K, Yoshida N, Kajiyama H, Shibata K, Yamamoto E, Kidokoro K, et al. Indoleamine 2,3-dioxygenase is a novel prognostic indicator for endometrial cancer. Br J Cancer 2006 Dec 4;95(11):1555–61.

[4] Ino K, Yamamoto E, Shibata K, Kajiyama H, Yoshida N, Terauchi M, et al. Inverse correlation between tumoral indoleamine 2,3-dioxygenase expression and tumor-infiltrating lymphocytes in endometrial cancer: its association with disease progression and survival. Clin Cancer Res 2008 Apr 15;14(8):2310–7.

[5] Okamoto A, Nikaido T, Ochiai K, Takakura S, Saito M, Aoki Y, et al. Indoleamine 2,3-dioxygenase serves as a marker of poor prognosis in gene expression profiles of serous ovarian cancer cells. Clin Cancer Res 2005 Aug 15;11(16):6030–9.

[6] Takao M, Okamoto A, Nikaido T, Urashima M, Takakura S, Saito M, et al. Increased synthesis of indoleamine-2,3-dioxygenase protein is positively associated with impaired survival in patients with serous-type, but not with other types of, ovarian cancer. Oncol Rep 2007 Jun;17(6):1333–9.

[7] Brandacher G, Perathoner A, Ladurner R, Schneeberger S, Obrist P, Winkler C, et al. Prognostic value of indoleamine 2,3-dioxygenase expression in colorectal cancer: effect on tumor-infiltrating T cells. Clin Cancer Res 2006 Feb 15;12(4):1144–51.

[8] Pan K, Wang H, Chen MS, Zhang HK, Weng DS, Zhou J, et al. Expression and prognosis role of indoleamine 2,3-dioxygenase in hepatocellular carcinoma. J Cancer Res Clin Oncol 2008 Nov;134(11):1247–53.

[9] Lob S, Konigsrainer A. Is IDO a key enzyme bridging the gap between tumor escape and tolerance induction? Langenbecks Arch Surg 2008 Nov;393(6):995–1003.

[10] Munn DH, Zhou M, Attwood JT, Bondarev I, Conway SJ, Marshall B, et al. Prevention of allogeneic fetal rejection by tryptophan catabolism. Science 1998 Aug 21;281(5380):1191–3.

[11] Brandacher G, Cakar F, Winkler C, Schneeberger S, Obrist P, Bosmuller C, et al. Non-invasive monitoring of kidney allograft rejection through IDO metabolism evaluation. Kidney Int 2007 Jan;71(1):60–7.

[12] Yoshida N, Ino K, Ishida Y, Kajiyama H, Yamamoto E, Shibata K, et al. Overexpression of indoleamine 2,3-dioxygenase in human endometrial carcinoma cells induces rapid tumor growth in a mouse xenograft model. Clin Cancer Res 2008 Nov 15;14(22):7251–9.

[13] Muller AJ, DuHadaway JB, Donover PS, Sutanto-Ward E, Prendergast GC. Inhibition of indoleamine 2,3-dioxygenase, an immunoregulatory target of the cancer suppression gene Bin1, potentiates cancer chemotherapy. Nat Med 2005 Mar;11(3):312–9.

[14] Qian F, Villella J, Wallace PK, Mhawech-Fauceglia P, Tario Jr JD, Andrews C, et al. Efficacy of levo-1-methyl tryptophan and dextro-1-methyl tryptophan in reversing indoleamine-2,3-dioxygenase-mediated arrest of T-cell proliferation in human epithelial ovarian cancer. Cancer Res 2009 Jul 1;69(13):5498–504.

[15] Metz R, Duhadaway JB, Kamasani U, Laury-Kleintop L, Muller AJ, Prendergast GC. Novel tryptophan catabolic enzyme IDO2 is the preferred biochemical target of the antitumor indoleamine 2,3-dioxygenase inhibitory compound D-1-methyl-tryptophan. Cancer Res 2007 Aug 1;67(15):7082–7.



Brad H. Nelson

*Trev and Joyce Deeley Research Centre, BC Cancer Agency, 2410 Lee Avenue, Victoria, BC, Canada V8R 6V5*

*E-mail address:* bnelson@bccancer.bc.ca.

OPEN ACCESS Freely available online

PLoS one

# Systematic Analysis of Immune Infiltrates in High-Grade Serous Ovarian Cancer Reveals CD20, FoxP3 and TIA-1 as Positive Prognostic Factors

**Katy Milne[1], Martin Köbel[2], Steven E. Kalloger[2], Rebecca O. Barnes[1], Dongxia Gao[3], C. Blake Gilks[2], Peter H. Watson[1,3,4], Brad H. Nelson[1,4,5]***

**1** Trev and Joyce Deeley Research Centre, BC Cancer Agency, Victoria, British Columbia, Canada, **2** Department of Anatomical Pathology, Vancouver General Hospital, Vancouver, British Columbia, Canada, **3** Department of Pathology, University of British Columbia, Vancouver, British Columbia, Canada, **4** Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia, Canada, **5** Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada

## Abstract

***Background:*** Tumor-infiltrating T cells are associated with survival in epithelial ovarian cancer (EOC), but their functional status is poorly understood, especially relative to the different risk categories and histological subtypes of EOC.

***Methodology/Principal Findings:*** Tissue microarrays containing high-grade serous, endometrioid, mucinous and clear cell tumors were analyzed immunohistochemically for the presence of lymphocytes, dendritic cells, neutrophils, macrophages, MHC class I and II, and various markers of activation and inflammation. In high-grade serous tumors from optimally debulked patients, positive associations were seen between intraepithelial cells expressing CD3, CD4, CD8, CD45RO, CD25, TIA-1, Granzyme B, FoxP3, CD20, and CD68, as well as expression of MHC class I and II by tumor cells. Disease-specific survival was positively associated with the markers CD8, CD3, FoxP3, TIA-1, CD20, MHC class I and class II. In other histological subtypes, immune infiltrates were less prevalent, and the only markers associated with survival were MHC class II (positive association in endometrioid cases) and myeloperoxidase (negative association in clear cell cases).

***Conclusions/Significance:*** Host immune responses to EOC vary widely according to histological subtype and the extent of residual disease. TIA-1, FoxP3 and CD20 emerge as new positive prognostic factors in high-grade serous EOC from optimally debulked patients.

**Citation:** Milne K, Köbel M, Kalloger SE, Barnes RO, Gao D, et al. (2009) Systematic Analysis of Immune Infiltrates in High-Grade Serous Ovarian Cancer Reveals CD20, FoxP3 and TIA-1 as Positive Prognostic Factors. PLoS ONE 4(7): e6412. doi:10.1371/journal.pone.0006412

**Editor:** Derya Unutmaz, New York University School of Medicine, United States of America

**Received** May 19, 2009; **Accepted** June 26, 2009; **Published** July 29, 2009



**Funding:** The study was funded by grants from the British Columbia Cancer Foundation, Vancouver General Hospital Foundation, and U.S. Dept. of Defense (OC000018). None of the sponsors had any role in the design and conduct of the study, in the collection, analysis, and interpretation of the data, or in the preparation, review, or approval of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: bnelson@bccancer.bc.ca

## Introduction

Ovarian cancer is the most deadly gynecologic cancer, affecting more than 190,000 women worldwide each year (International Agency for Research on Cancer). Delayed diagnosis and the presence of widely disseminated disease account for the high mortality associated with the disease. Additionally, while a large percentage of patients initially respond well to cytoreductive surgery and standard chemotherapy, the disease usually recurs within 2-5 years as residual tumor cells develop resistance to chemotherapy [1,2]. Although prognosis is often poor, numerous favorable prognostic indicators have been described, including early stage, low grade and optimal surgical debulking [3,4].

Several recent studies have analyzed the influence of host immunity on disease prognosis. Tumor-infiltrating CD3+ T cells are strongly associated with favorable prognosis, specifically when CD3+ cells are localized within tumor epithelium [5-9]. These findings have been extended to the CD8+ T cell subset in particular [10-17], suggesting that cytotoxic T lymphocytes (CTLs) play an important role in the antitumor immune response. Accordingly, other factors associated with CTL responses are also positively associated with survival, including interferon-γ (IFN- γ) [18,19], the IFN- γ receptor [20], interferon regulatory factor (IRF)-1 [21], IL-18 [22], TNF-α [23], MHC class I [24-26], and MHC class I antigen processing machinery [17].

In contrast to CD8+ T cells, several studies have indicated that tumor-infiltrating CD25+FoxP3+ T cells (referred to as regulatory T cells or Tregs) are associated with decreased survival [10,27-29]. Tregs have the ability to suppress proliferation, cytokine production, and cytolytic activity of CD4+ and CD8+ T cells by mechanisms involving cell-to-cell contact and the release of cytokines such as TGF-β [30,31]. Tregs can also induce an immunosuppressive phenotype in other cell types such as macrophages [32]. Although Tregs have been associated with poor prognosis in many cancers, several exceptions have recently been reported. Leffers et. al. found that FoxP3+ infiltrates in advanced stage EOC were associated with increased survival [14]. Similar findings have been reported in colorectal cancer [33] and

lymphoma [34-36]. Furthermore, in murine models, FoxP3+ cells can play a positive role in anti-tumor and anti-viral immunity [37,38]. The precise role of regulatory T cells in cancer outcomes warrants further consideration given that several groups are attempting to enhance tumor immunity by depleting FoxP3+ Tregs from cancer patients [39-44], including EOC patients [45].

In addition to Tregs, other cell types reportedly play an immunosuppressive role in EOC. For example, plasmacytoid dendritic cells contribute to immunosuppression by promoting the development or recruitment of interleukin-10-producing CD4+ and CD8+ regulatory T cells [46,47]. Myeloid dendritic cells (MDCs) impair T cell immunity by expressing B7-H1, a ligand for the inhibitory receptor PD-1 found on T cells [48]. Monocytes and macrophages in the EOC microenvironment can be polarized toward a so-called M2 phenotype, which is typified by the expression of IL-10, TGF-b and scavenger receptors and is thought to promote tumor progression [49,50,51]. Under the influence of IL-6 and IL-10, macrophages in EOC can also express B7-H4, which inhibits T cell proliferation [52]. Macrophages also produce CCL22, which promotes Treg recruitment to the tumor environment [32]. Finally, expression of the inflammatory mediator COX-2 in tumor epithelium has been associated with reduced lymphocyte infiltration and poor prognosis in EOC [13,53].

With the advent of tumor tissue microarray (TMA) technology, a large number of retrospective studies have investigated the relationship between tumor-infiltrating immune cells and prognosis in EOC and other cancers. However, most studies focus on one or a few markers, such that associations between different immunological factors may be missed. Additionally most studies fail to address the different histological subtypes of EOC, which are now recognized to behave as distinct diseases [54]. As a result, there are inconsistencies and unresolved issues in the literature concerning the prognostic significance of different immune cell infiltrates. To address this, we analyzed several large series of EOC tumors, including high-grade serous, endometrioid, clear cell and mucinous subtypes, for the presence of various immune cell infiltrates and inflammatory markers. Our results reveal that high-grade serous tumors have a distinct immunological profile that is strongly associated with patient survival.

## Results

### Intraepithelial T cells and associated functional markers in high-grade serous EOC

We initially investigated the relationship between immune infiltrates and survival in a cohort of 199 high-grade serous EOC patients. We chose to first focus on high-grade serous cases, as the other histological subtypes exhibit distinct biological and clinical properties that are potentially confounding [54,55]. This initial cohort was restricted to patients who had undergone optimal cytoreduction (i.e., without evidence of macroscopic residual disease). Patient characteristics are shown in Table 1.

The tumors in this initial cohort had been previously assessed by immunohistochemistry (IHC) for a variety of lymphocyte markers, including CD3, CD4, CD8, CD20 and Granzyme B [12]. Intraepithelial lymphocytes (i.e., lymphocytes within the epithelial component of the tumor) were scored as either present (i.e. one or more intraepithelial lymphocytes present in at least one of two 0.6 mm cores) or absent. We re-analyzed this data focusing exclusively on high-grade serous cases. We found that 83.2% (163/196) of evaluable high-grade serous tumors were positive for intraepithelial CD3+ T cells, whereas CD4+ and CD8+ intraepithelial cells were found in 53.4% (103/193) and 84.0% (163/194) of evaluable tumors, respectively (Fig. 1A&B and data not shown). CD4+ and CD8+ cellular infiltrates showed a strong positive association ($p<0.0001$). Table 2 shows statistical associations for these and all other markers studied in this initial cohort.

**Table 1.** Clinical characteristics of the optimally debulked high-grade serous EOC patient cohort.

| | |
|---|---|
| Age at surgery (years) | |
| Mean | 61.00 |
| Std dev | 11.48 |
| Range | 37.59-85.96 |
| Median | 60.08 |
| * Overall Survival (years) | |
| Mean | 5.59 |
| Std dev | 3.47 |
| Range | 0.4-17.4 |
| Median | 4.91 |
| Silverberg Grade | |
| 1 | 0 |
| 2 | 56 |
| 3 | 143 |
| Unknown | 0 |
| Stage | |
| I | 49 |
| II | 85 |
| III | 65 |
| IV | 0 |
| Unknown | 0 |
| Total number of evaluable patients | 199 |

*There were no deaths due to causes other than ovarian cancer, therefore disease-specific and overall survival were equivalent.

doi:10.1371/journal.pone.0006412.t001

While the above markers indicate which lymphocyte subsets are present in tumors, they do not reveal their activation state. To address this issue, we analyzed tumors for expression of CD45RO, OX40 and CD25, which are expressed by activated T cells [56,57]. Using the same scoring criteria as above, 70.6% (132/187) of tumors were positive for intraepithelial CD45RO+ cells, and 49.7% (96/193) were positive for intraepithelial CD25+ cells (Fig. 1C&E). By contrast, only 7.0% (11/158) of tumors were positive for intraepithelial OX40+ cells (Fig. 1D). In pair-wise comparisons, CD45RO, CD25 and OX40 were all positively associated (Table 2). Moreover, CD45RO and CD25 were both associated with the presence of CD3+, CD4+ and CD8+ cells. OX40 showed a similar trend, but this did not reach statistical significance, likely due to the low number of positive cases.

To investigate the differentiation state of tumor-infiltrating T cells, tissues were analyzed for intraepithelial cells expressing TIA-1 and Granzyme B, which are markers of CD8+ cytotoxic T cells and NK cells [58-60]. A majority of tumors (66.5%, 127/191) were positive for intraepithelial TIA-1+ cells (Fig. 1F), and about half of tumors (45.6%, 88/193) were positive for intraepithelial Granzyme B+ cells (Fig. 1G). There was a highly significant association between TIA-1+ and Granzyme B+ cells ($p<0.0001$). Moreover, in pair-wise comparisons, TIA-1+ and Granzyme B+ cells were each associated with the activation markers CD45RO, CD25 and OX40. Finally, TIA-1+ and Granzyme B+ cells were each associated with the presence of CD3+, CD4+ and CD8+ cells (Table 2). To examine whether TIA-1 and Granzyme B

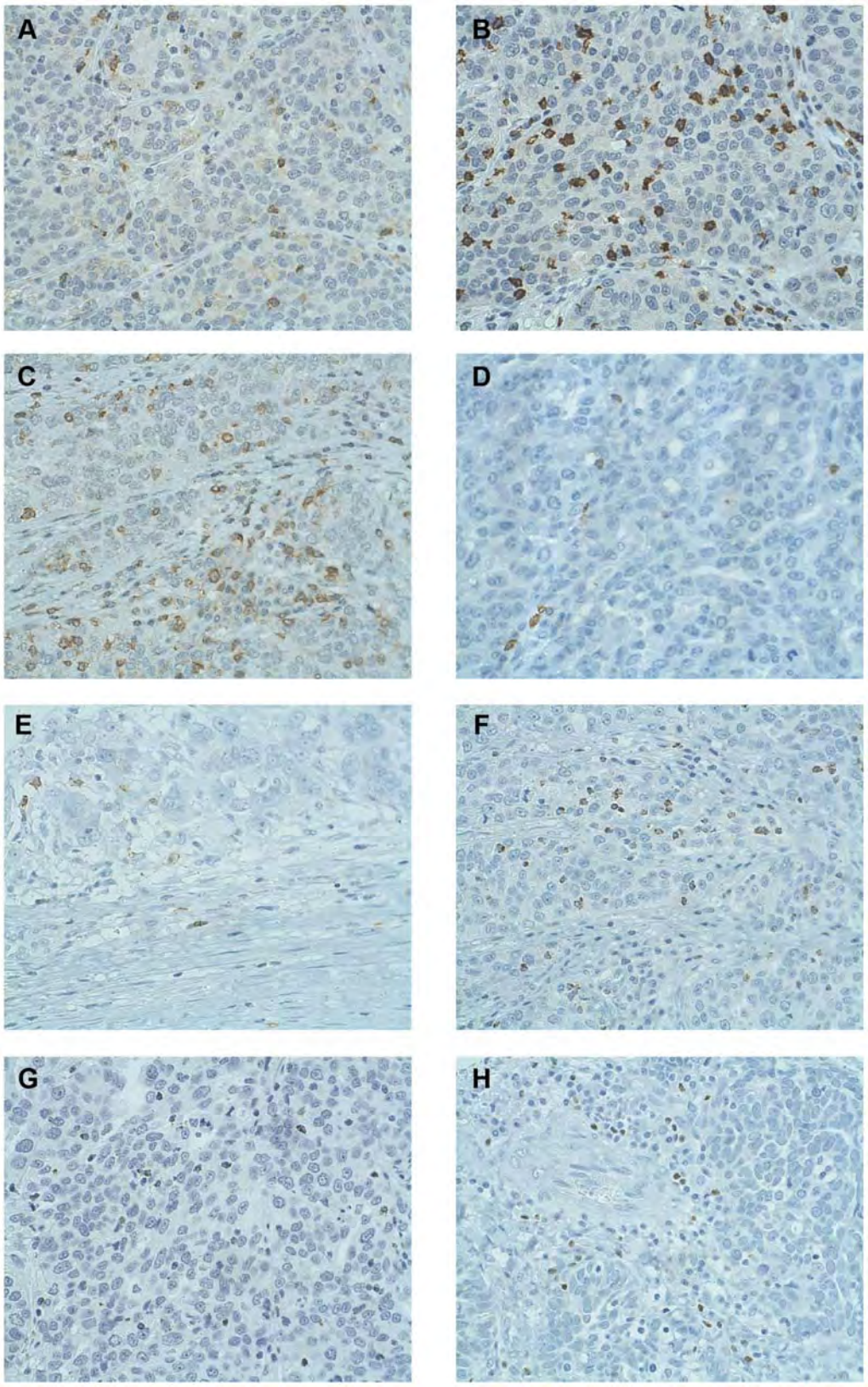



**Figure 1. Immunohistochemical analysis of high-grade serous EOC tumors showing infiltrates expressing markers of T cell differentiation and activation.** (A) CD4, (B) CD8, (C) CD45RO, (D) OX40, (E) CD25, (F) TIA-1, (G) Granzyme B, and (H) FoxP3. 40X objective.
doi:10.1371/journal.pone.0006412.g001

**Table 2.** *p*-values for Chi-square tests of associations between immunohistochemical markers in the optimally debulked high-grade serous EOC cohort.

| | CD3 | CD8 | CD4 | CD45R0 | CD25 | OX40 | TIA-1 | GrB[1] |
|---|---|---|---|---|---|---|---|---|
| **CD3** | | <0.0001 | <0.0001 | <0.0001 | <0.0001 | 0.16 | <0.0001 | <0.0001 |
| **CD8** | <0.0001 | | <0.0001 | <0.0001 | <0.0001 | 0.15 | <0.0001 | <0.0001 |
| **CD4** | <0.0001 | <0.0001 | | 0.0021 | <0.0001 | 0.049 | 0.0013 | 0.0001 |
| **CD45RO** | <0.0001 | <0.0001 | 0.0021 | | <0.0001 | 0.031 | <0.0001 | <0.0001 |
| **CD25** | <0.0001 | <0.0001 | <0.0001 | <0.0001 | | 0.0040 | <0.0001 | <0.0001 |
| **OX40** | 0.16 | 0.15 | 0.049 | 0.031 | 0.0040 | | 0.012 | 0.044 |
| **TIA-1** | <0.0001 | <0.0001 | 0.0013 | <0.0001 | <0.0001 | 0.012 | | <0.0001 |
| **GrB[1]** | <0.0001 | <0.0001 | 0.0001 | <0.0001 | <0.0001 | 0.044 | <0.0001 | |
| **FoxP3** | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | 0.47 | <0.0001 | <0.0001 |
| **MHC I[2]** | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | 0.18 | <0.0001 | 0.0001 |
| **MHC II[3]** | 0.0078 | 0.0009 | 0.013 | 0.0004 | 0.0001 | 0.38 | 0.0006 | 0.016 |
| **CD20** | <0.0001 | <0.0001 | 0.0006 | 0.0036 | 0.0006 | 0.13 | <0.0001 | <0.0001 |
| **CD1a** | 0.11 | 0.096 | 0.046 | 0.21 | 0.27 | 0.030 | 0.31 | 0.48 |
| **CD68** | <0.0001 | 0.0001 | 0.0028 | 0.0019 | 0.0006 | 0.23 | 0.0005 | 0.0007 |
| **MPO[4]** | 0.89 | 0.53 | 0.034 | 0.90 | 0.78 | 0.33 | 0.67 | 0.49 |
| **COX-2[5]** | 0.75 | 0.64 | 0.55 | 0.64 | 0.74 | 0.76 | 0.42 | 0.70 |

| | FoxP3 | MHC I[2] | MHC II[3] | CD20 | CD1a | CD68 | MPO[4] | COX2[5] |
|---|---|---|---|---|---|---|---|---|
| **CD3** | <0.0001 | <0.0001 | 0.0078 | <0.0001 | 0.11 | <0.0001 | 0.89 | 0.75 |
| **CD8** | <0.0001 | <0.0001 | 0.0009 | <0.0001 | 0.096 | 0.0001 | 0.53 | 0.64 |
| **CD4** | <0.0001 | <0.0001 | 0.013 | 0.0006 | 0.046 | 0.0028 | 0.034 | 0.55 |
| **CD45RO** | <0.0001 | <0.0001 | 0.0004 | 0.0036 | 0.21 | 0.0019 | 0.90 | 0.64 |
| **CD25** | <0.0001 | <0.0001 | 0.0001 | 0.0006 | 0.27 | 0.0006 | 0.78 | 0.74 |
| **OX40** | 0.47 | 0.18 | 0.38 | 0.13 | 0.030 | 0.23 | 0.33 | 0.76 |
| **TIA-1** | <0.0001 | <0.0001 | 0.0006 | <0.0001 | 0.31 | 0.0005 | 0.67 | 0.42 |
| **GrB[1]** | <0.0001 | 0.0001 | 0.016 | <0.0001 | 0.48 | 0.0007 | 0.49 | 0.70 |
| **FoxP3** | | <0.0001 | <0.0001 | 0.0009 | 0.35 | 0.0001 | 0.42 | 0.82 |
| **MHC I[2]** | <0.0001 | | <0.0001 | 0.0015 | 0.13 | 0.035 | 0.61 | 0.17 |
| **MHC II[3]** | <0.0001 | <0.0001 | | 0.023 | 0.11 | 0.22 | 0.39 | 0.30 |
| **CD20** | 0.0009 | 0.0015 | 0.023 | | 0.87 | 0.029 | 0.60 | 0.13 |
| **CD1a** | 0.35 | 0.13 | 0.11 | 0.87 | | 0.52 | 0.046 | 0.84 |
| **CD68** | 0.0001 | 0.035 | 0.22 | 0.029 | 0.52 | | 0.43 | 0.30 |
| **MPO[4]** | 0.42 | 0.61 | 0.39 | 0.60 | 0.046 | 0.43 | | 0.51 |
| **COX-2[5]** | 0.82 | 0.17 | 0.30 | 0.13 | 0.84 | 0.30 | 0.51 | |

[1]GrB = Granzyme B.
[2]MHC I = MHC class I.
[3]MHC II = MHC class II.
[4]MPO = myeloperoxidase.
[5]COX-2 = Cyclooxygenase-2.
doi:10.1371/journal.pone.0006412.t002

expression could be due to the presence of NK or NKT cells, we examined tumors for the NK cell markers CD56 and CD57. For both markers, there were either few or no infiltrates at all within the tumor epithelium (data not shown), indicating that the TIA-1+ and Granzyme B+ infiltrates were most likely T cells.

Finally, tumors were analyzed for the presence of intraepithelial cells expressing FoxP3, which in humans is a marker of regulatory T cells and activated T cells [61,62]. About half of tumors (52.9%, 100/189) were positive for intraepithelial FoxP3+ cells (Fig. 1H). There was a strong association between FoxP3+ and CD25+ cells ($p<0.0001$), and FoxP3+ and CD25+ cells were each strongly associated with CD4+ cells ($p<0.0001$ for both markers). Thus, consistent with previous reports [10,14,27,28], a significant proportion of tumors contained intraepithelial infiltrates with markers characteristic of Tregs (CD4+,CD25+, and FoxP3+).

## MHC class I and II in high-grade serous EOC

We analyzed tumor cells for expression of MHC class I and II using a four-point scale (negative, focal [<10%], patchy [10-50%] or diffuse [>50%]). All evaluable tumors (185/185) expressed MHC class I to some degree (i.e., focal, patchy or diffuse), indicating they could theoretically present antigen to CD8+ T cells. For

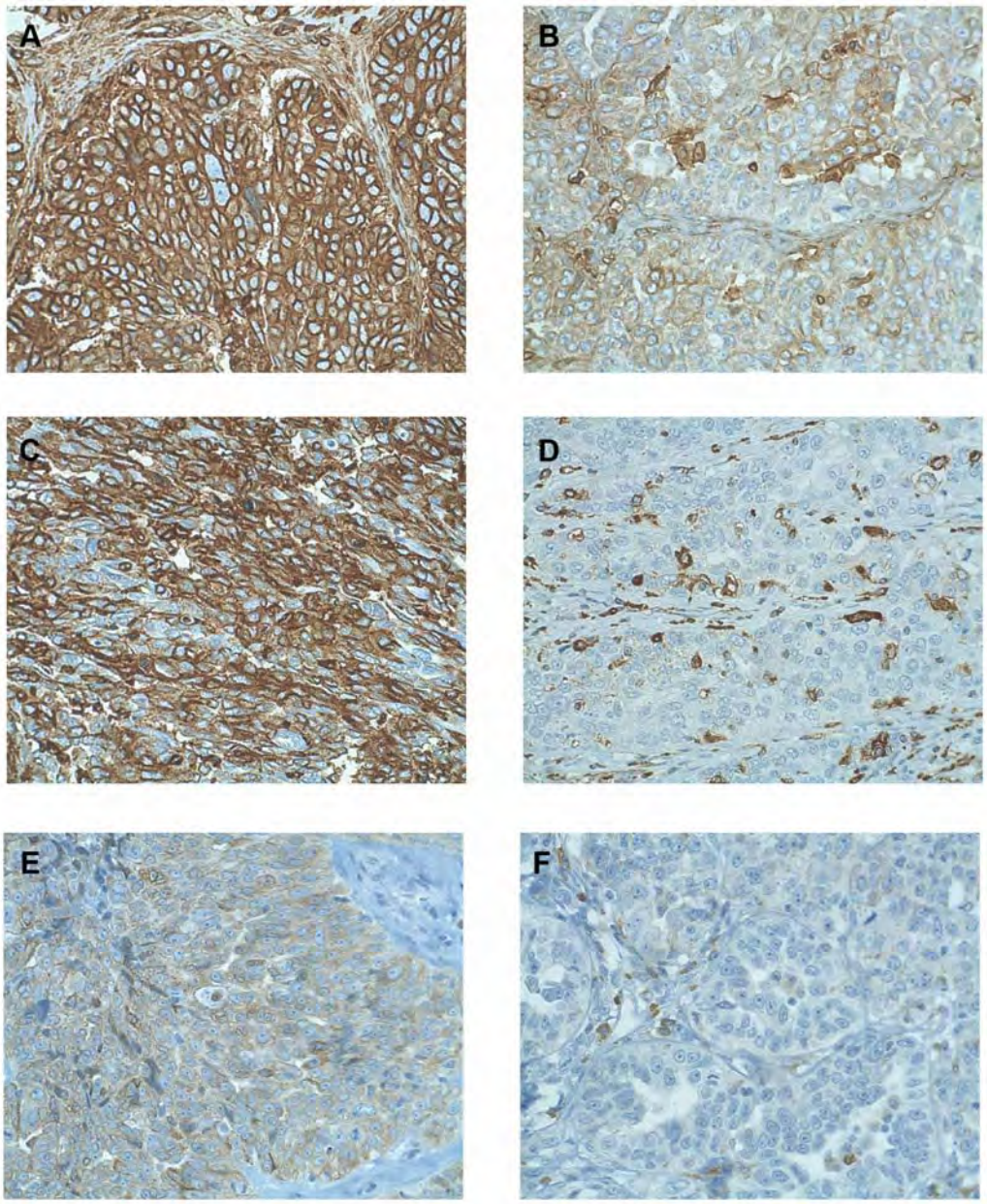



**Figure 2. Immunohistochemical analysis of high-grade serous EOC tumors showing (A,B) high and low expression of MHC class I, (C,D) high and low expression of MHC class II, and (E,F) high and low expression of COX-2.** 40X objective.
doi:10.1371/journal.pone.0006412.g002

statistical analyses, only the highest category (diffuse, >50%) was considered positive (Fig. 2A&B). Using this threshold, 85.4% (158/185) of tumors were positive for MHC class I. MHC class I was positively associated with all three T cell subsets (CD3, CD4, and CD8), the activation markers CD45RO and CD25, and the differentiation markers TIA-1, Granzyme B and FoxP3 (Table 2).

A large majority of tumors (86.5%, 166/192) expressed MHC class II to some degree (i.e., focal, patchy or diffuse), indicating they could theoretically present antigen to CD4+ T cells. As with MHC class I, only the highest category (diffuse, >50%) was considered positive for statistical analyses (Fig. 2 C&D). Using this threshold, 41.1% (79/192) of tumors were positive for MHC class II. MHC class II was strongly associated with MHC class I ($p<0.0001$). Accordingly, MHC class II was positively associated with all three T cell subsets (CD3, CD4, and CD8), the activation markers CD45RO and CD25, and the differentiation markers TIA-1, Granzyme B and FoxP3 (Table 2). Similar to the results for MHC class I, the expression of MHC class II in tumor epithelium was positively associated with various T cell markers, including CD3, CD4, CD8, CD45RO, TIA-1, Granzyme B, CD25 and FoxP3 (Table 2).

## Intraepithelial B cells in high-grade serous EOC

Tissues were stained with an antibody to CD20, which is expressed by B cells from the naïve to memory stages of differentiation [63]. Intraepithelial CD20+ cells were found in 41.9% (83/198) of evaluable tumors (Fig. 3A). CD20+ infiltrates were strongly associated with all three T cell subsets (CD3, CD4, and CD8); the activation markers CD45RO and CD25; the differentiation markers TIA-1, Granzyme B and FoxP3; and both MHC class I and II (Table 2).

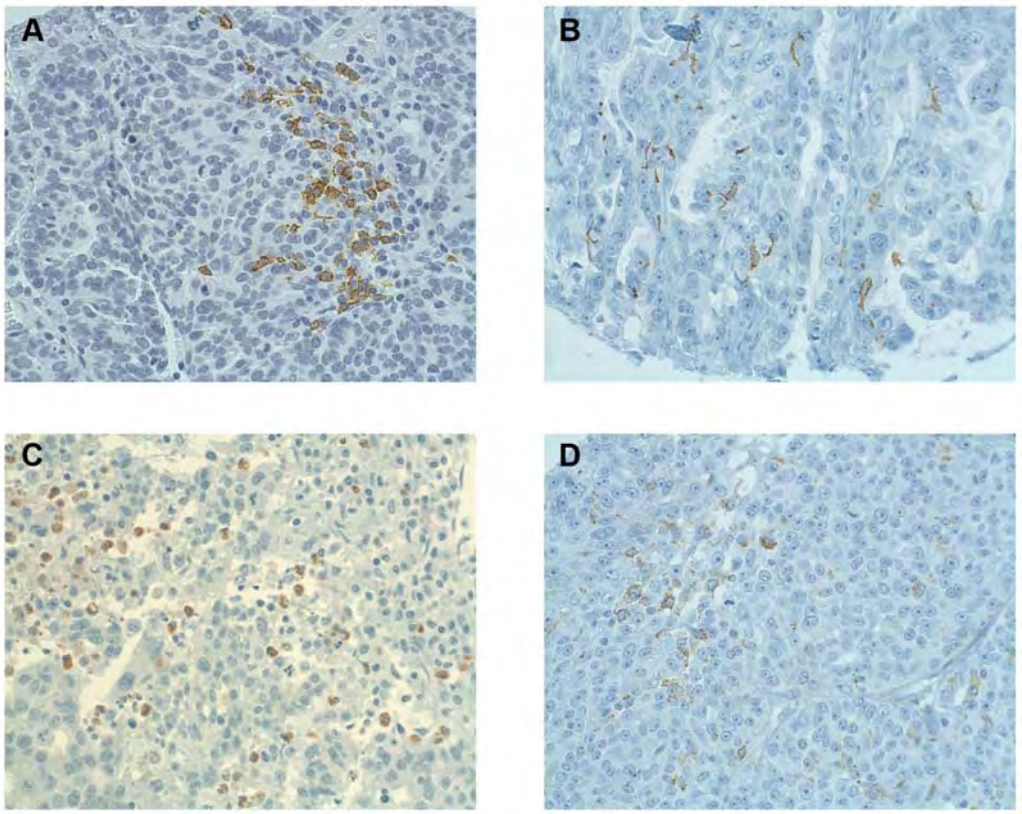



**Figure 3. Immunohistochemical analysis of high-grade serous EOC tumors showing infiltrates expressing (A) CD20 (B cells), (B) CD1a (immature DCs), (C) Myeloperoxidase (granulocytes), and (D) CD68 (macrophages).** 40X objective.
doi:10.1371/journal.pone.0006412.g003

## Intraepithelial dendritic cells, granulocytes and macrophages in high-grade serous EOC

Tumors were analyzed for the presence of immature and mature dendritic cells by staining for CD1a and CD208, respectively. A minority of tumors (13.4%, 23/172) contained intraepithelial CD1a+ cells (Fig. 3B). No significant association with any of the intraepithelial lymphocyte markers (CD3, CD8 or CD20), activation markers (CD45RO or CD25), differentiation markers (TIA-1, Granzyme B or FoxP3) or MHC class I or II (Table 2) was seen, potentially due to the low number of CD1a+ cells. In contrast to CD1a, none of the tumors scored positive for intraepithelial CD208+ cells. Parallel analysis of tonsil tissue revealed the presence of many CD208+ cells, thereby validating the IHC procedure.

About half of tumors (54.7%, 87/159) contained intraepithelial cells expressing the macrophage marker CD68 (Fig. 3D). CD68 was positively associated with several lymphocyte markers (CD3, CD8 and CD20), activation markers (CD45RO and CD25), differentiation markers (TIA-1, Granzyme B and FoxP3) and MHC class I (Table 2). To assess the presence of granulocytes, the TMA was stained for myeloperoxidase. Twenty four percent (37/154) of tumors contained myeloperoxidase-expressing cells (Fig. 3C), however these showed no significant associations with other markers, with the exception of CD4 ($p = 0.034$).

The COX-2 enzyme has been associated with inferior survival in EOC when expressed in the epithelial component of the tumor [53]. Therefore, tumors were scored for expression of COX-2 in the epithelial component using a four-point scale (negative, equivocal [0-1%], patchy [1-50%] or diffuse [>50%]) (Fig. 2E&F). Two-thirds of tumors (66.5%, 111/167) were positive for COX-2 (i.e., patchy or diffuse staining). In contrast to reports in ovarian, cervical, and other cancers, [13,64,65], the expression of COX-2 was not significantly associated with any of the immune infiltrates studied (Table 2).

## Associations between immune infiltrates and patient survival in high-grade serous EOC

Kaplan-Meier analysis was performed to assess the association between various immune infiltrates and disease-specific survival (DSS). Consistent with prior reports [5-7,10-17], intraepithelial CD3+ and CD8+ cells were associated with increased DSS ($p = 0.0009$ and 0.0008 respectively) (Fig. 4A&B). Intraepithelial CD4+ cells showed a trend towards increased DSS, but this was not statistically significant (Fig. 4C). The NK cell markers CD56 and CD57 showed no association with DSS (data not shown).

Intriguingly, intraepithelial CD20+ cells were associated with increased DSS ($p = 0.0033$) (Fig. 4E). Furthermore, the combination of CD8+ and CD20+ infiltrates was associated with significantly increased DSS over tumors that contained CD8+ infiltrates but not CD20+ infiltrates (median 4432 days vs. 2279 days, $p = 0.0115$) (data not shown).

In contrast to lymphocyte markers, the dendritic cell marker CD1a showed no association with DSS, possibly due to the low number of tumors containing CD1a+ cells. Likewise, the markers CD68, COX-2 and myeloperoxidase showed no association with DSS (Fig. 4F and data not shown).

Given the association between CD8+ T cells and DSS, we evaluated other canonical features of active CTL responses. DSS was positively associated with intraepithelial TIA-1+ cells ($p = 0.0003$), as well as expression of MHC class I and II by

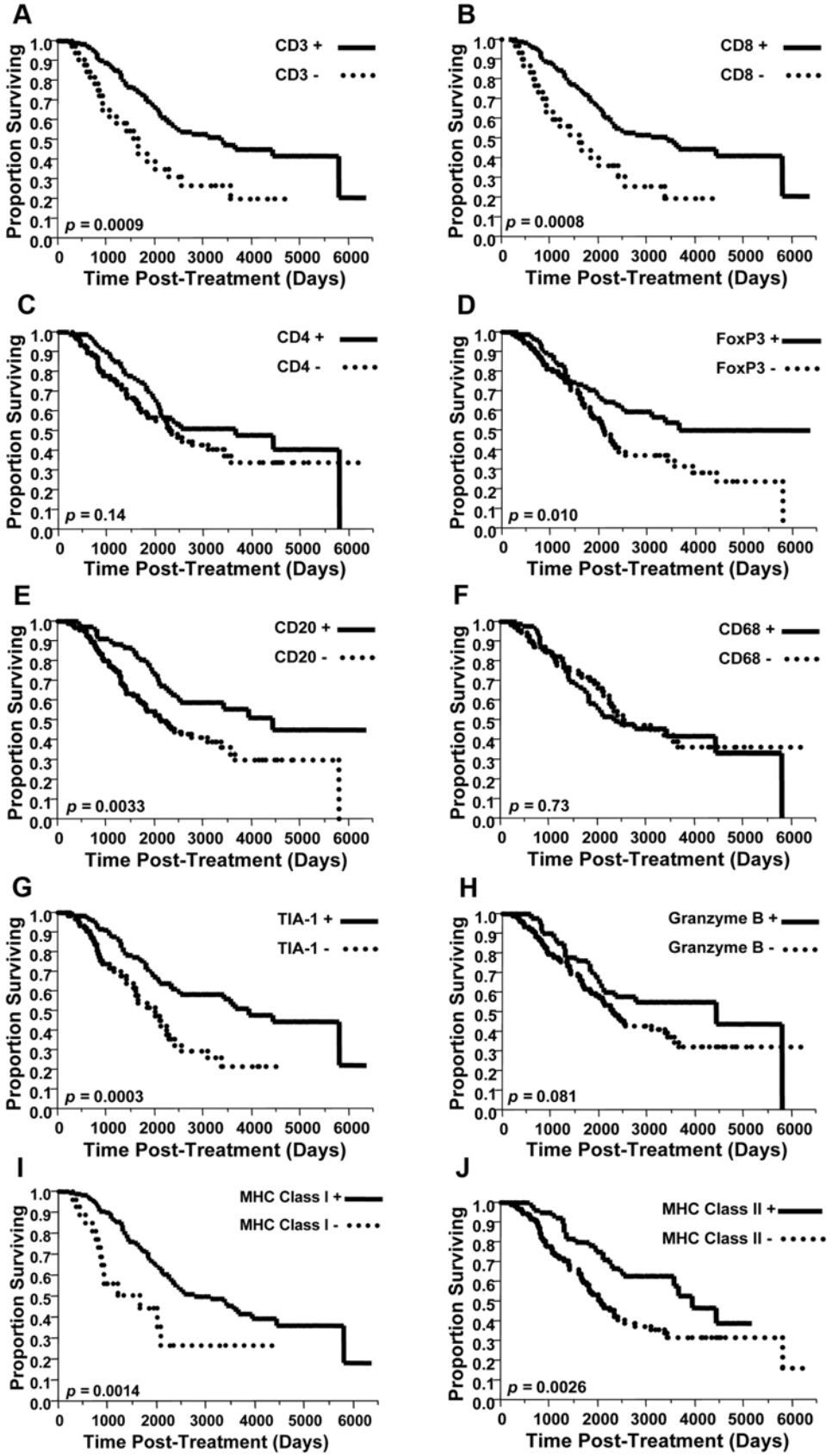



**Figure 4. Immune infiltrates and survival in ovarian cancer.** Kaplan-Meier curves showing disease-specific survival for patients scored as positive or negative for (A) CD3, (B) CD8, (C) CD4, (D) FoxP3, (E) CD20, (F) CD68, (G) TIA-1, (H) Granzyme B, (I) MHC Class I and (J) MHC Class II. Data were derived from optimally debulked patients with high-grade serous EOC.
doi:10.1371/journal.pone.0006412.g004

tumor cells ($p = 0.0014$ and 0.0026 respectively) (Fig. 4G,I,&J). Tumors that contained both CD8+ and TIA-1+ infiltrates were associated with increased DSS compared to CD8+ TIA-1-negative tumors ($p = 0.0025$). Several other T cell markers, including Granzyme B, CD45RO and CD25, showed trends toward increased DSS but did not reach statistical significance (Fig. 4H and data not shown). OX-40 showed no apparent trend or association with DSS, possibly due to low numbers of positive cases (data not shown).

In apparent contrast to reports that regulatory T cells are associated with poor prognosis, the presence of intraepithelial FoxP3+ cells was associated with increased DSS ($p = 0.010$) (Fig. 4D). Moreover, tumors that were triply positive for intraepithelial CD4+, FoxP3+ and CD25+ cells showed a trend towards increased survival, although this fell short of statistical significance ($p = 0.059$). Likewise, tumors positive for both intraepithelial CD8+ and FoxP3+ cells showed a trend toward increased DSS compared to tumors that were positive for CD8+ cells but negative for FoxP3+ cells; however, this trend did not reach statistical significance ($p = 0.052$). Thus, by multiple analyses, tumor-infiltrating FoxP3+ cells showed a trend or statistically significant association with increased DSS.

## The association between immune infiltrates and survival is dependent on the extent of residual disease

T cell infiltrates are reportedly more prevalent in patients with optimal versus suboptimal cytoreduction [5,66]. To investigate whether this was true for other lymphocyte markers, we analyzed an additional cohort of 220 high-grade serous cases from patients known to have macroscopic residual disease following primary cytoreductive surgery. We focused on CD8+ infiltrates, as well as the three novel prognostic markers from the preceding analysis (i.e., FoxP3, TIA-1 and CD20). Compared to the optimally debulked patient cohort, patients with macroscopic residual disease had a significantly lower prevalence of CD8+ (58.5%), FoxP3+ (20.2%), TIA-1+ (39.5%) and CD20+ (16.3%) infiltrates ($p<0.0001$ for all markers). In Kaplan-Meier analysis of these four markers, only CD8+ infiltrates had a significant association with survival ($p = 0.0044$) in patients with macroscopic residual disease (data not shown).

## The association between immune infiltrates and survival is dependent on histological subtype

The preceding results were based exclusively on high-grade serous EOC cases. To assess the association between immune infiltrates and DSS in other histological subtypes of EOC, we performed the same analyses using an additional 288 EOC tumors of the following histological subtypes: mucinous ($n = 31$), endometrioid ($n = 125$) and clear cell ($n = 132$). These additional tumor specimens were from a previously described cohort of optimally debulked patients [12].

In general, immune infiltrates were less prevalent in the other histological subtypes compared to the high-grade serous cases discussed previously. This was true for all lymphocyte markers studied (i.e., CD3, CD8, CD4, CD45RO, CD25, FoxP3, TIA-1, Granzyme B, and CD20) (Fig. 5). The difference was most striking for the markers FoxP3, CD25 and CD20. After the high-grade serous cases, the next highest frequency of immune infiltrates was seen in the endometrioid subtype (Fig. 5).

We examined the association between immune infiltrates and DSS in the endometrioid and clear cell subtypes; the number of mucinous cases was too small to perform robust statistical analysis. For endometrioid cases, the only significant association found was between MHC class II expression and increased DSS ($p = 0.039$) (data not shown). For clear cell cases, the only significant association found was between the presence of myeloperoxidase-positive infiltrates and decreased DSS ($p = 0.040$, data not shown). Thus, the relationship between immune infiltrates and survival differs greatly between histological subtypes of EOC.

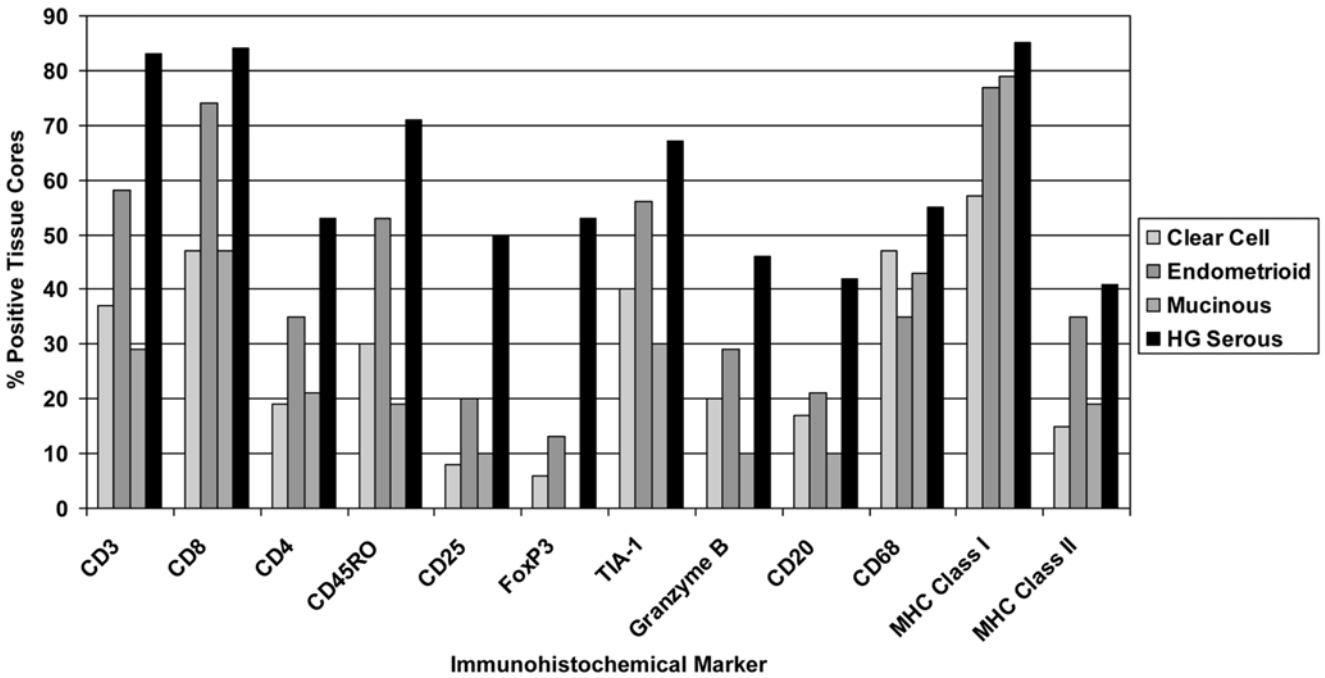



**Figure 5. Prevalence of immune infiltrates and other markers across different histologic subtypes of EOC.** Bars indicate the percentage of tumors scoring positive for intrapithelial cells expressing CD3, CD8, CD4, CD45RO, CD25, FoxP3, TIA-1, Granzyme B, CD20 and CD68. Expression of MHC class I and II by tumor epithelium is also shown. Data were derived from optimally debulked patients.
doi:10.1371/journal.pone.0006412.g005

## Discussion

We systematically examined the relationship between immune infiltrates and patient survival in three large EOC series. In accord with Clarke et al. [12], we found that high-grade serous tumors have a distinct immunological profile compared to the endometrioid, clear cell and mucinous subtypes. Furthermore, we found that immune infiltrates were generally more prevalent in tumors from patients with optimal cytoreduction. FoxP3, TIA-1 and CD20 emerged as novel immunological markers associated with increased patient survival. Our results highlight the importance of histological subtype in the immunobiology of EOC, which may have important implications for the immunotherapy of this family of diseases.

Intraepithelial lymphocytes (i.e., cells expressing CD3, CD4, CD8, FoxP3 or CD20) were more prevalent in high-grade serous cases, followed by endometrioid cases. Moreover, intraepithelial lymphocytes were more prevalent in tumors from optimally debulked patients compared to patients with macroscopic residual disease. A number of biological features of tumors appear to influence the density of lymphocytic infiltrates. (a) T cell infiltrates are positively associated with expression of MHC class I and II by tumor cells (Table 2), as well as MHC class I antigen processing machinery [15-17,67], suggesting that antigen presentation may be an important determinant of T cell infiltration. (b) In accord with this notion, tumors with loss or mutation of the BRCA1 or p53 genes have an increased density of tumor-infiltrating T cells [12,66]. This suggests that defective DNA repair and the ensuing genomic instability in tumors may lead to the generation of neo-antigens that trigger host T cell responses. (c) Signaling molecules also play a role, as the density of tumor-infiltrating T cells is negatively associated with expression of VEGF, B7-H1/PD-L1 and endothelin B receptor by tumors [5,11,68] and positively associated with expression of the chemokines CXCL9, CCL21, CCL22, CCL2 and CCL5 [5,28,69]. (d) Finally, two groups have reported gene expression profiles that correlate with the presence of tumor-infiltrating T cells in EOC [12,70]. These latter studies confirm some of the above associations (e.g., MHC class I and II, beta 2 microglobulin, TAP1 and 2) and identify new factors associated with T cell infiltrates (e.g., IL-15, IL-32 and numerous interferon-induced genes). Presumably one or more of the above factors accounts for the observed enrichment of tumor-infiltrating lymphocytes in high-grade serous and optimally cytoreduced cases.

Although the association between intraepithelial CD8+ T cells and increased survival in EOC is a highly reproducible finding [10-17], relatively little is known about the functional phenotype of these CD8+ T cells. Several lines of evidence suggest a classic cytolytic response underlies favorable outcomes. For example, others have reported positive associations between survival and intratumoral expression of IFN- $\gamma$ [18,19], the IFN- $\gamma$ receptor [20], as well as numerous interferon-responsive genes such as MHC class I [24-26], MHC class I antigen processing machinery [17], MHC class II [15,16], and IRF-1 [21]. IL-18 [22] and TNF-$\alpha$ [23] also appear to be important components of the T cell response, as both cytokines are positively associated with survival. We examined two components of cytolytic granules, Granzyme B and TIA-1, both of which showed an association with CD8+ T cell infiltrates. Of these two markers, only TIA-1 showed a statistically significant association with survival in high-grade serous cases (Fig. 4). TIA-1+ cells have also been described in medullary breast cancer [71,72] and melanoma [73], where they are associated with favorable prognostic features. By contrast, tumor-infiltrating TIA-1+ cells are associated with decreased survival in lymphoma [74-78]. Interestingly, TIA-1 is not simply a marker of cytolytic granules; it is an RNA binding protein involved in post-transcriptional mRNA regulation [79]. It remains to be determined whether the association between intraepithelial TIA-1+ cells and survival in EOC is due to the role of this protein in cytolytic granule function or mRNA regulation.

Treg infiltrates have previously been associated with decreased survival in ovarian EOC [10,27,28]. However, in the present study and one other recent report [14], FoxP3+ infiltrates were associated with increased survival. These seemingly contradictory findings may be attributable to several factors. First, not all studies take into consideration the histological subtypes of EOC, or the extent of residual disease; in the present study, FoxP3+ cells were only associated with survival in high-grade serous tumors from optimally debulked patients. Second, a variety of antibodies have been used to detect FoxP3, which can lead to discordant results [80]. Third, different scoring criteria may be used. For example, the precise intratumoral location of Tregs is an important determinant of prognosis in gastric cancer [81]. Fourth, the molecular markers used to define Tregs differ between studies. Although FoxP3 is still regarded as the most reliable marker of Tregs in human cancer [82,83], it can also be expressed by epithelial tumor cells [84-86] and *in vitro* activated CD4+ and CD8+ T cells [87-95]. For these reasons, some studies include CD25 as a second marker of Tregs [10,28]. However, like FoxP3, CD25 is potentially expressed by effector T cells, so it is not clear that dual staining for FoxP3 and CD25 more accurately identifies Tregs [89,96]. Other characteristics of Tregs include high expression of GITR and CTLA-4 and low expression of CD127 and CD49d and [97,98], however these markers are technically difficult to assess on paraffin-embedded TMAs.

These technical considerations notwithstanding, there is mounting evidence that tumor-infiltrating FoxP3+ cells are associated with a favorable prognosis in EOC, colorectal cancer, head and neck cancer, and lymphoma [14,33,-36,99-102]. How might FoxP3+ T cells promote favorable outcomes? In the present study, FoxP3+ cells were strongly associated with other effector T cells, and similar results have been reported in melanoma [103]. Thus, FoxP3+ cells may simply be an indicator of a strong CD8+ T cell response, which might outweigh any immunosuppressive effects of FoxP3+ cells. Alternatively, subsets of human FoxP3+ T cells have recently been shown to have a pro-inflammatory, IL-17-producing phenotype [104-106]. Indeed, CD4+ T cells can be skewed toward this so-called Th17 phenotype by exposure to TGF-$\beta$ in combination with IL-6, IL-1 or IL-23 [107-109]. These factors are present in the EOC tumor environment [8], and accordingly, Th17 cells have been reported in EOC [110-112]. Thus, the association between FoxP3+ cells and increased survival could potentially reflect an underlying Th17-like anti-tumor response. Clearly, more work is required to determine the extent to which FoxP3+ T cells in EOC represent Tregs versus Th17 or other effector T cells.

The observation that intraepithelial CD20+ infiltrates are associated with increased DSS is a novel finding in EOC. Dong et. al. reported that B cells in ascites were associated with shorter survival in EOC [113], however their study focused on B cells in peritoneal and pleural effusions collected after chemotherapy, which by definition constitutes a poor outcome cohort. Indeed, in the present study, intraepithelial CD20+ B cells showed no association with survival in patients with high-risk, suboptimally debulked disease. Tumor-infiltrating CD20+ B cells are a hallmark of medullary breast cancer and have been proposed to mediate a favorable prognosis [114,115]. Moreover, the presence of a B cell transcriptional signature in node-negative breast cancer is

associated with increased survival [116]. B cell infiltrates in breast cancer represent clonally expanded populations, express somatically hypermutated IgG molecules, and recognize target antigens such as ganglioside D3 and surface-translocated actin [114,115,117-120]. It is unclear how tumor-infiltrating B cells promote favorable outcomes in cancer. In theory, their actions could be mediated by secreted antibodies, which can promote the opsonization of tumor antigens, complement-mediated destruction of tumor cells, or antibody-dependent cellular cytotoxicity. Apart from producing antibodies, B cells can also present antigen to both CD4+ and CD8+ T cells [121-131]. In this regard, it is noteworthy that ovarian tumors show low numbers of CD1a+ dendritic cells; perhaps CD20+ B cells serve as alternative antigen presenting cells in the tumor environment. This latter idea fits well with the observed co-localization of tumor-infiltrating B cells and CD8+ T cells in EOC, as well as in medullary breast cancer, non-small cell lung cancer and cervical cancer [72,132-134].

The presence of macrophages has been associated with poor prognosis in various human cancers [135,136]. However, in accord with a prior report by Shah et. al. [66], we found no association between CD68+ infiltrates and survival in EOC. Importantly, however, CD68 is not a perfect marker of macrophages, as it is also expressed by dendritic cells and some non-myeloid cells [137]. Furthermore, CD68 does not distinguish between macrophages polarized towards the pro-inflammatory (M1) or tumor-promoting (M2) phenotypes. M1 macrophages have the capacity to kill tumor cells, whereas M2 macrophages promote tissue repair and angiogenesis [136]. Similarly, an immunosuppressive subpopulation of macrophages has been described in EOC based on expression of the signaling molecule B7-H4 [52]. Thus, additional functional markers may be required to fully define the role of macrophages in the immunobiology of EOC.

While this study focused on the relationship between immune infiltrates and prognosis after standard treatments, the results may also inform the design of immunotherapies for EOC. First, our findings suggest that high-grade serous tumors may be especially sensitive to T cell responses. Second, our data indicates that, in patients with residual disease, the influence of T cells may be overwhelmed by other factors. Third, the positive association between intraepithelial FoxP3+ cells and survival reported here and previously [14] prompts a reconsideration of strategies to deplete regulatory T cells from EOC patients. And fourth, the association between intraepithelial CD20+ cells and survival suggests the humoral immune response may play an important role in anti-tumor immunity that could be exploited therapeutically in parallel with CD8+ T cell responses.

## Materials and Methods

### Study subjects

All specimens and clinical data were obtained with informed written consent under protocols approved by the Research Ethics Board of the BC Cancer Agency and the University of British Columbia. The main cohort used for this study consisted of 199 women with high-grade serous ovarian cancer seen at the BC Cancer Agency from 1984 to 2000 (OvCaRe Ovarian Tumour Bank, Vancouver, BC, Canada). Tumor tissue was obtained at the time of primary surgery prior to any other treatment. Patients had no macroscopic residual disease following surgical debulking. All patients underwent standard treatment consisting of surgery followed by standard platinum-based chemotherapy. Table 1 shows the general clinical characteristics of the 199-case cohort. We also analyzed a second cohort of mucinous (n = 31), endometrioid (n = 125) and clear cell (n = 132) EOC cases. Patients in this cohort were also diagnosed from 1984 to 2000, were optimally cytoreduced, and received platinum-based chemotherapy. Finally, we analyzed a third cohort of 220 high-grade serous EOC patients categorized as extreme risk due to the presence of residual macroscopic disease. Patients in this cohort were treated from 1996 to 2000 and received platinum-based chemotherapy.

### Tumor specimens

Tumor tissue was obtained during primary cytoreductive surgery, fixed in 10% neutral buffered formalin, processed using standard procedures and embedded in paraffin. A tissue microarray (TMA) was constructed by taking duplicate 0.6 mm cores from representative regions of each tumor block after review of hematoxylin- and eosin-stained sections by a pathologist. TMAs were assembled using a Pathology Devices tissue arrayer (Westminster, MD).

### Immunohistochemistry

Immunohistochemistry for CD20, CD3, CD4, CD8 and Granzyme B was performed as described in Clarke et al.[12]. The remaining unstained slides were received at the Trev and Joyce Deeley Research Centre where immunohistochemistry was performed for CD45RO, TIA-1, FoxP3, CD25, OX-40, CD56, CD57, CD1a, CD208, myeloperoxidase, CD68, COX-2, MHC Class I and MHC Class II. Following deparaffinization, the slides were placed in a Ventana Discovery XT autostainer (Ventana, Tucson, AZ) for immunohistochemical staining. Ventana's standard CC1 protocol was used for antigen retrieval. Primary antibodies are listed in Table 3.

TMAs were incubated with primary antibodies for 60 minutes at room temperature, and the appropriate cross-adsorbed, biotinylated secondary antibody (Jackson Immunoresearch, West Grove, PA) was applied for 32 minutes. Bound antibodies were detected using the DABMap kit (Ventana), counterstained with hematoxylin (Ventana), and coverslipped manually with Cytoseal-60 (Richard Allan, Kalamazoo, MI).

### Histopathological analysis

Immunostained TMAs were examined by a pathologist and scored using a variety of methods depending on the marker studied. For CD20, CD8, CD4, CD45RO, TIA-1, Granzyme B, CD25, OX40, CD1a, CD56, CD57, and myeloperoxidase, only cells residing within the epithelial compartment of the tumor were counted. FoxP3 was similarly scored within the epithelium but a stromal score was also obtained. Tumors were scored as 0 (no cells), 1 (1-5 cells), 2 (6-19 cells) or 3 (20+ cells); results were binarized as positive (IHC score 1, 2, or 3) or negative (IHC score 0). CD3 was scored as 0 (no cells present), 1 (cells present in stroma only), 2 (cells present in the epithelial compartment) or 3 (cells present in both the epithelial and stromal regions of the tumor); scores of 0 and 1 were reported as negative while a score of 2 or 3 was reported as positive. CD68 was scored as 0 (no cells present), 1 (luminal or stromal cells), 2 (scattered, $<$20 intraepithelial cells), or 3 ($>$20 intraepithelial cells), results were binarized in the same manner as CD3. COX-2 was scored as 0 (negative), 1 (equivocal, 0-1%), 2 (patchy, $>$1% to 50%) or 3 (diffuse, $>$50%) scores of 0 and 1 were reported as negative and scores of 2 or 3 were reported as positive. For MHC class I and II, samples were scored as 0 (negative), 1 (focal, $<$10%), 2 (patchy, 10-50%) or 3 (diffuse, $>$50%). Scores of 0, 1 or 2 were reported as negative and scores of 3 were reported as positive.

**Table 3.** Primary antibodies used for immunohistochemistry.

| Antigen | Clone | Supplier | Source | Concentration |
|---|---|---|---|---|
| CD3 | Polyclonal | Cell Marque | Rabbit | 1/300 |
| CD8 | C8/144C | DAKO | Rabbit | 1/50 |
| CD4 | 4B12 | Novocastra | Mouse | 1/50 |
| Granzyme B | GrB-7 | DAKO | Mouse | 1/25 |
| CD45RO | UCHL-1 | Lab Vision | Mouse | 1/300 |
| TIA-1 | TIA-1 | Abcam | Mouse | 1/50 |
| FoxP3 | eBio7979 | eBioscience | Mouse | 1/50 |
| CD25 | 4C9 | Lab Vision | Mouse | 1/40 |
| OX-40 (CD134) | ACT35 | BD Pharmingen | Mouse | 1/50 |
| CD20 | L26 | DAKO | Mouse | 1/250 |
| CD56 | 123C3.D5 | Lab Vision | Mouse | 1/50 |
| CD57 | NK1 | Lab Vision | Mouse | 1/200 |
| CD1a | O10 | Lab Vision | Mouse | 1/50 |
| CD208 | 1010E1.01 | Imgenix | Rat | 1/50 |
| Myeloperoxidase | Polyclonal, Catalogue # RB-373 | Lab Vision | Rabbit | 1/200 |
| CD68 | PG-M1 | Lab Vision | Mouse | 1/50 |
| Cyclooxygenase-2 (COX-2) | SP21 | Cell Marque | Rabbit | 1/10 |
| MHC class I (A, B, C) | EMR8-5 | MBL | Mouse | 1/500 |
| MHC class II (DR, DP & DQ) | CR3/43 | Affinty BioReagents | Mouse | 1/50 |

doi:10.1371/journal.pone.0006412.t003

### Statistical analysis

Statistical analysis was performed using JMP statistical software (v7.0) (SAS Institute, Cary, NC). Univariate analysis was carried out using the Chi-Squared statistic. The log-rank test was used to compare Kaplan-Meier curves. *p*-values less than 0.05 were considered significant.

## Acknowledgments

We thank Jennifer Santos, Erika Mehl, Rob Grigor, and members of the Deeley Research Centre for assistance and advice.

## Author Contributions

Conceived and designed the experiments: KM BN. Performed the experiments: KM. Analyzed the data: KM MK SEK DG CBG. Contributed reagents/materials/analysis tools: SEK RB. Wrote the paper: KM BN. Performed histological scoring: MK DG CBG. Contributed suggestions/revisions to paper: MK SEK RB CBG PHW. Ethics acquisition: RB.

## References

1. Ozols RF, Bundy BN, Greer BE, Fowler JM, Clarke-Pearson D, et al. (2003) Phase III trial of carboplatin and paclitaxel compared with cisplatin and paclitaxel in patients with optimally resected stage III ovarian cancer: a Gynecologic Oncology Group study. J Clin Oncol 21: 3194–3200.
2. du Bois A, Luck HJ, Meier W, Adams HP, Mobus V, et al. (2003) A randomized clinical trial of cisplatin/paclitaxel versus carboplatin/paclitaxel as first-line treatment of ovarian cancer. J Natl Cancer Inst 95: 1320–1329.
3. Holschneider CH, Berek JS (2000) Ovarian cancer: epidemiology, biology, and prognostic factors. Semin Surg Oncol 19: 3–10.
4. Ozols RF (2000) Management of advanced ovarian cancer consensus summary. Advanced Ovarian Cancer Consensus Faculty. Semin Oncol 27: 47–49.
5. Zhang L, Conejo-Garcia JR, Katsaros D, Gimotty PA, Massobrio M, et al. (2003) Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer. N Engl J Med 348: 203–213.
6. Raspollini MR, Castiglione F, Rossi Degl'innocenti D, Amunni G, Villanucci A, et al. (2005) Tumour-infiltrating gamma/delta T-lymphocytes are correlated with a brief disease-free interval in advanced ovarian serous carcinoma. Ann Oncol 16: 590–596.
7. Tomsova M, Melichar B, Sedlakova I, Steiner I (2008) Prognostic significance of CD3+ tumor-infiltrating lymphocytes in ovarian carcinoma. Gynecol Oncol 108: 415–420.
8. Nelson BH (2008) The impact of T-cell immunity on ovarian cancer outcomes. Immunol Rev 222: 101–116.
9. Gimotty PA, Zhang L, Alagkiozidis I, Cadungog M, Adams S, et al. (2007) Immune prognostic factors in ovarian cancer: lessons from translational research. Dis Markers 23: 445–452.
10. Sato E, Olson SH, Ahn J, Bundy B, Nishikawa H, et al. (2005) Intraepithelial CD8+ tumor-infiltrating lymphocytes and a high CD8+/regulatory T cell ratio are associated with favorable prognosis in ovarian cancer. Proc Natl Acad Sci U S A 102: 18538–18543.
11. Hamanishi J, Mandai M, Iwasaki M, Okazaki T, Tanaka Y, et al. (2007) Programmed cell death 1 ligand 1 and tumor-infiltrating CD8+ T lymphocytes are prognostic factors of human ovarian cancer. Proc Natl Acad Sci U S A 104: 3360–3365.
12. Clarke B, Tinker AV, Lee CH, Subramanian S, van de Rijn M, et al. (2009) Intraepithelial T cells and prognosis in ovarian carcinoma: novel associations with stage, tumor type, and BRCA1 loss. Mod Pathol 22: 393–402.
13. Liu M, Matsumura N, Mandai M, Li K, Yagi H, et al. (2009) Classification using hierarchical clustering of tumor-infiltrating immune cells identifies poor prognostic ovarian cancers with high levels of COX expression. Mod Pathol 22: 373–384.
14. Leffers N, Gooden MJ, de Jong RA, Hoogeboom BN, ten Hoor KA, et al. (2009) Prognostic significance of tumor-infiltrating T-lymphocytes in primary and metastatic lesions of advanced stage ovarian cancer. Cancer Immunol Immunother 58: 449–459.
15. Callahan MJ, Nagymanyoki Z, Bonome T, Johnson ME, Litkouhi B, et al. (2008) Increased HLA-DMB expression in the tumor epithelium is associated with increased CTL infiltration and improved prognosis in advanced-stage serous ovarian cancer. Clin Cancer Res 14: 7667–7673.
16. Matsushita N, Ghazizadeh M, Konishi H, Araki T (2003) Association of ovarian tumor epithelium coexpressing HLA-DR and CA-125 antigens with tumor infiltrating cytotoxic T lymphocytes. J Nippon Med Sch 70: 40–44.
17. Han LY, Fletcher MS, Urbauer DL, Mueller P, Landen CN, et al. (2008) HLA Class I Antigen Processing Machinery Component Expression and Intratu-

moral T-Cell Infiltrate as Independent Prognostic Markers in Ovarian Carcinoma. Clin Cancer Res 14: 3372–3379.
18. Marth C, Fiegl H, Zeimet AG, Muller-Holzner E, Deibl M, et al. (2004) Interferon-gamma expression is an independent prognostic factor in ovarian cancer. Am J Obstet Gynecol 191: 1598–1605.
19. Kusuda T, Shigemasa K, Arihiro K, Fujii T, Nagai N, et al. (2005) Relative expression levels of Th1 and Th2 cytokine mRNA are independent prognostic factors in patients with ovarian cancer. Oncol Rep 13: 1153–1158.
20. Duncan TJ, Rolland P, Deen S, Scott IV, Liu DT, et al. (2007) Loss of IFN gamma receptor is an independent prognostic factor in ovarian cancer. Clin Cancer Res 13: 4139–4145.
21. Zeimet AG, Reimer D, Wolf D, Fiegl H, Concin N, et al. (2008) Intratumoral interferon regulatory factor (IRF)-1 but not IRF-2 is of relevance in predicting patient outcome in ovarian cancer. Int J Cancer.
22. Akahiro J, Konno R, Ito K, Okamura K, Yaegashi N (2004) Impact of serum interleukin-18 level as a prognostic indicator in patients with epithelial ovarian carcinoma. Int J Clin Oncol 9: 42–46.
23. Bamias A, Tsiatas ML, Kafantari E, Liakou C, Rodolakis A, et al. (2007) Significant differences of lymphocytes isolated from ascites of patients with ovarian cancer compared to blood and tumor lymphocytes. Association of CD3(+)CD56(+) cells with platinum resistance. Gynecol Oncol.
24. Rolland P, Deen S, Scott I, Durrant L, Spendlove I (2007) Human leukocyte antigen class I antigen expression is an independent prognostic factor in ovarian cancer. Clin Cancer Res 13: 3591–3596.
25. Moore DH, Fowler WC Jr, Olafsson K (1990) Class I histocompatibility antigen expression: a prognostic factor for aneuploid ovarian cancers. 38: 458–461.
26. Leffers N, Lambeck AJ, de Graeff P, Bijlsma AY, Daemen T, et al. (2008) Survival of ovarian cancer patients overexpressing the tumour antigen p53 is diminished in case of MHC class I down-regulation. Gynecol Oncol 110: 365–373.
27. Wolf D, Wolf AM, Rumpold H, Fiegl H, Zeimet AG, et al. (2005) The expression of the regulatory T cell-specific forkhead box transcription factor FoxP3 is associated with poor prognosis in ovarian cancer. Clin Cancer Res 11: 8326–8331.
28. Curiel TJ, Coukos G, Zou L, Alvarez X, Cheng P, et al. (2004) Specific recruitment of regulatory T cells in ovarian carcinoma fosters immune privilege and predicts reduced survival. Nat Med 10: 942–949.
29. Woo EY, Chu CS, Goletz TJ, Schlienger K, Yeh H, et al. (2001) Regulatory CD4(+)CD25(+) T cells in tumors from patients with early-stage non-small cell lung cancer and late-stage ovarian cancer. Cancer Res 61: 4766–4772.
30. Miyara M, Sakaguchi S (2007) Natural regulatory T cells: mechanisms of suppression. Trends Mol Med 13: 108–116.
31. Piccirillo CA (2008) Regulatory T cells in health and disease. Cytokine 43: 395–401.
32. Kryczek I, Wei S, Zhu G, Myers L, Mottram P, et al. (2007) Relationship between B7-H4, regulatory T cells, and patient outcome in human ovarian carcinoma. Cancer Res 67: 8900–8905.
33. Salama P, Phillips M, Grieu F, Morris M, Zeps N, et al. (2009) Tumor-infiltrating FOXP3+ T regulatory cells show strong prognostic significance in colorectal cancer. J Clin Oncol 27: 186–192.
34. de Jong D, Koster A, Hagenbeek A, Raemaekers J, Veldhuizen D, et al. (2009) Impact of the tumor microenvironment on prognosis in follicular lymphoma is dependent on specific treatment protocols. Haematologica 94: 70–77.
35. Ke X, Wang J, Li L, Chen IH, Wang H, et al. (2008) Roles of CD4+CD25(high) FOXP3+ Tregs in lymphomas and tumors are complex. Front Biosci 13: 3986–4001.
36. Carreras J, Lopez-Guillermo A, Fox BC, Colomo L, Martinez A, et al. (2006) High numbers of tumor-infiltrating FOXP3-positive regulatory T cells are associated with improved overall survival in follicular lymphoma. Blood 108: 2957–2964.
37. Erdman SE, Sohn JJ, Rao VP, Nambiar PR, Ge Z, et al. (2005) CD4+CD25+ regulatory lymphocytes induce regression of intestinal tumors in ApcMin/+ mice. Cancer Res 65: 3998–4004.
38. Lund JM, Hsing L, Pham TT, Rudensky AY (2008) Coordination of early protective immunity to viral infection by regulatory T cells. Science 320: 1220–1224.
39. Mahnke K, Schonfeld K, Fondel S, Ring S, Karakhanova S, et al. (2007) Depletion of CD4(+)CD25(+) human regulatory T cells in vivo: Kinetics of Treg depletion and alterations in immune functions in vivo and in vitro. Int J Cancer 120: 2723–2733.
40. Morse MA, Hobeika AC, Osada T, Serra D, Niedzwiecki D, et al. (2008) Depletion of human regulatory T cells specifically enhances antigen-specific immune responses to cancer vaccines. Blood 112: 610–618.
41. Powell DJ Jr, Attia P, Ghetie V, Schindler J, Vitetta ES, et al. (2008) Partial reduction of human FOXP3+ CD4 T cells in vivo after CD25-directed recombinant immunotoxin administration. J Immunother 31: 189–198.
42. Powell DJ Jr, Felipe-Silva A, Merino MJ, Ahmadzadeh M, Allen T, et al. (2007) Administration of a CD25-directed immunotoxin, LMB-2, to patients with metastatic melanoma induces a selective partial reduction in regulatory T cells in vivo. J Immunol 179: 4919–4928.
43. Dannull J, Su Z, Rizzieri D, Yang BK, Coleman D, et al. (2005) Enhancement of vaccine-mediated antitumor immunity in cancer patients after depletion of regulatory T cells. J Clin Invest 115: 3623–3633.
44. Rasku MA, Clem AL, Telang S, Taft B, Gettings K, et al. (2008) Transient T cell depletion causes regression of melanoma metastases. J Transl Med 6: 12.
45. Barnett B, Kryczek I, Cheng P, Zou W, Curiel TJ (2005) Regulatory T cells in ovarian cancer: biology and therapeutic potential. Am J Reprod Immunol 54: 369–377.
46. Wei S, Kryczek I, Zou L, Daniel B, Cheng P, et al. (2005) Plasmacytoid dendritic cells induce CD8+ regulatory T cells in human ovarian carcinoma. Cancer Res 65: 5020–5026.
47. Zou W, Machelon V, Coulomb-L'Hermin A, Borvak J, Nome F, et al. (2001) Stromal-derived factor-1 in human tumors recruits and alters the function of plasmacytoid precursor dendritic cells. Nat Med 7: 1339–1346.
48. Curiel TJ, Wei S, Dong H, Alvarez X, Cheng P, et al. (2003) Blockade of B7-H1 improves myeloid dendritic cell-mediated antitumor immunity. Nat Med 9: 562–567.
49. Duluc D, Delneste Y, Tan F, Moles MP, Grimaud L, et al. (2007) Tumor-associated leukemia inhibitory factor and IL-6 skew monocyte differentiation into tumor-associated-macrophage-like cells. Blood.
50. Hagemann T, Wilson J, Burke F, Kulbe H, Li NF, et al. (2006) Ovarian cancer cells polarize macrophages toward a tumor-associated phenotype. J Immunol 176: 5023–5032.
51. Loercher AE, Nash MA, Kavanagh JJ, Platsoucas CD, Freedman RS (1999) Identification of an IL-10-producing HLA-DR-negative monocyte subset in the malignant ascites of patients with ovarian carcinoma that inhibits cytokine protein expression and proliferation of autologous T cells. J Immunol 163: 6251–6260.
52. Kryczek I, Zou L, Rodriguez P, Zhu G, Wei S, et al. (2006) B7-H4 expression identifies a novel suppressive macrophage population in human ovarian carcinoma. J Exp Med 203: 871–881.
53. Denkert C, Kobel M, Pest S, Koch I, Berger S, et al. (2002) Expression of cyclooxygenase 2 is an independent prognostic factor in human ovarian carcinoma. Am J Pathol 160: 893–903.
54. Kobel M, Kalloger SE, Boyd N, McKinney S, Mehl E, et al. (2008) Ovarian carcinoma subtypes are different diseases: implications for biomarker studies. PLoS Med 5: e232.
55. Gilks CB, Ionescu DN, Kalloger SE, Kobel M, Irving J, et al. (2008) Tumor cell type can be reproducibly diagnosed and is of independent prognostic significance in patients with maximally debulked ovarian carcinoma. Human Pathology In press.
56. Brenchley JM, Douek DC, Ambrozak DR, Chatterji M, Betts MR, et al. (2002) Expansion of activated human naive T-cells precedes effector function. Clin Exp Immunol 130: 432–440.
57. Weinberg AD (2002) OX40: targeted immunotherapy–implications for tempering autoimmunity and enhancing vaccines. Trends Immunol 23: 102–109.
58. Pipkin ME, Lieberman J (2007) Delivering the kiss of death: progress on understanding how perforin works. Curr Opin Immunol 19: 301–308.
59. Kawakami A, Tian Q, Duan X, Streuli M, Schlossman SF, et al. (1992) Identification and functional characterization of a TIA-1-related nucleolysin. Proc Natl Acad Sci U S A 89: 8681–8685.
60. Kanavaros P, Boulland ML, Petit B, Arnulf B, Gaulard P (2000) Expression of cytotoxic proteins in peripheral T-cell and natural killer-cell (NK) lymphomas: association with extranodal site, NK or Tgammadelta phenotype, anaplastic morphology and CD30 expression. Leuk Lymphoma 38: 317–326.
61. Ziegler SF (2007) FOXP3: not just for regulatory T cells anymore. Eur J Immunol 37: 21–23.
62. Zou W (2006) Regulatory T cells, tumour immunity and immunotherapy. Nat Rev Immunol 6: 295–307.
63. Edwards JC, Cambridge G (2006) B-cell targeting in rheumatoid arthritis and other autoimmune diseases. Nat Rev Immunol 6: 394–403.
64. Ahmadi M, Emery DC, Morgan DJ (2008) Prevention of both direct and cross-priming of antitumor CD8+ T-cell responses following overproduction of prostaglandin E2 by tumor cells in vivo. Cancer Res 68: 7520–7529.
65. Lang S, Tiwari S, Andratschke M, Loehr I, Lauffer L, et al. (2007) Immune restoration in head and neck cancer patients after in vivo COX-2 inhibition. Cancer Immunol Immunother 56: 1645–1652.
66. Shah CA, Allison KH, Garcia RL, Gray HJ, Goff BA, et al. (2008) Intratumoral T cells, tumor-associated macrophages, and regulatory T cells: association with p53 mutations, circulating tumor DNA and survival in women with ovarian cancer. Gynecol Oncol 109: 215–219.
67. Kooi S, Zhang HZ, Patenia R, Edwards CL, Platsoucas CD, et al. (1996) HLA class I expression on human ovarian carcinoma cells correlates with T-cell infiltration in vivo and T-cell expansion in vitro in low concentrations of recombinant interleukin-2. 174: 116–128.
68. Buckanovich RJ, Facciabene A, Kim S, Benencia F, Sasaroli D, et al. (2008) Endothelin B receptor mediates the endothelial barrier to T cell homing to tumors and disables immune therapy. Nat Med 14: 28–36.
69. Negus RP, Stamp GW, Hadley J, Balkwill FR (1997) Quantitative assessment of the leukocyte infiltrate in ovarian cancer and its relationship to the expression of C-C chemokines. Am J Pathol 150: 1723–1734.
70. Tothill RW, Tinker AV, George J, Brown R, Fox SB, et al. (2008) Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. Clin Cancer Res 14: 5198–5208.

71. Kuroda H, Tamaru J, Sakamoto G, Ohnisi K, Itoyama S (2005) Immunophenotype of lymphocytic infiltration in medullary carcinoma of the breast. Virchows Arch 446: 10–14.
72. Yakirevich E, Izhak OB, Rennert G, Kovacs ZG, Resnick MB (1999) Cytotoxic phenotype of tumor infiltrating lymphocytes in medullary carcinoma of the breast. Mod Pathol 12: 1050–1056.
73. Lyle S, Salhany KE, Elder DE, Yakirevich E, Izhak OB, et al. (2000) TIA-1 positive tumor-infiltrating lymphocytes in nevi and melanomas.
74. Hasselblom S, Sigurdadottir M, Hansson U, Nilsson-Ehle H, Ridell B, et al. (2007) The number of tumour-infiltrating TIA-1+ cytotoxic T cells but not FOXP3+ regulatory T cells predicts outcome in diffuse large B-cell lymphoma. Br J Haematol 137: 364–373.
75. Chetaille B, Bertucci F, Finetti P, Esterni B, Stamatoullas A, et al. (2009) Molecular profiling of classical Hodgkin lymphoma tissues uncovers variations in the tumor microenvironment and correlations with EBV infection and outcome.
76. Asano N, Oshiro A, Matsuo K, Kagami Y, Ishida F, et al. (2006) Prognostic significance of T-cell or cytotoxic molecules phenotype in classical Hodgkin's lymphoma: a clinicopathologic study.
77. Alvaro-Naranjo T, Lejeune M, Salvado-Usach MT, Bosch-Princep R, Reverter-Branchat G, et al. (2005) Tumor-infiltrating cells as a prognostic factor in Hodgkin's lymphoma: a quantitative tissue microarray study in a large retrospective cohort of 267 patients.
78. Alvaro T, Lejeune M, Salvado MT, Bosch R, Garcia JF, et al. (2005) Outcome in Hodgkin's lymphoma can be predicted from the presence of accompanying cytotoxic and regulatory T cells.
79. Anderson P, Phillips K, Stoecklin G, Kedersha N (2004) Post-transcriptional regulation of proinflammatory proteins. J Leukoc Biol 76: 42–47.
80. Woo YL, Sterling J, Crawford R, van der Burg SH, Coleman N, et al. (2008) FOXP3 immunohistochemistry on formalin-fixed paraffin-embedded tissue: poor correlation between different antibodies. J Clin Pathol 61: 969–971.
81. Mizukami Y, Kono K, Kawaguchi Y, Akaike H, Kamimura K, et al. (2008) Localisation pattern of Foxp3+ regulatory T cells is associated with clinical behaviour in gastric cancer. Br J Cancer 98: 148–153.
82. Ahmadzadeh M, Felipe-Silva A, Heemskerk B, Powell DJ Jr. Wunderlich JR, et al. (2008) FOXP3 expression accurately defines the population of intratumoral regulatory T cells that selectively accumulate in metastatic melanoma lesions. Blood 112: 4953–4960.
83. Kryczek I, Liu R, Wang G, Wu K, Shu X, et al. (2009) FOXP3 defines regulatory T cells in human tumor and autoimmune disease. Cancer Res 69: 3995–4000.
84. Karanikas V, Speletas M, Zamanakou M, Kalala F, Loules G, et al. (2008) Foxp3 expression in human cancer cells. J Transl Med 6: 19.
85. Hinz S, Pagerols-Raluy L, Oberg HH, Ammerpohl O, Grussel S, et al. (2007) Foxp3 expression in pancreatic carcinoma cells as a novel mechanism of immune evasion in cancer. Cancer Res 67: 8344–8350.
86. Ebert LM, Tan BS, Browning J, Svobodova S, Russell SE, et al. (2008) The regulatory T cell-associated transcription factor FoxP3 is expressed by tumor cells. Cancer Res 68: 3001–3009.
87. Walker MR, Kasprowicz DJ, Gersuk VH, Benard A, Van Landeghen M, et al. (2003) Induction of FoxP3 and acquisition of T regulatory activity by stimulated human CD4+CD25- T cells. J Clin Invest 112: 1437–1443.
88. Morgan ME, van Bilsen JH, Bakker AM, Heemskerk B, Schilham MW, et al. (2005) Expression of FOXP3 mRNA is not confined to CD4+CD25+ T regulatory cells in humans. Hum Immunol 66: 13–20.
89. Roncador G, Brown PJ, Maestre L, Hue S, Martinez-Torrecuadrada JL, et al. (2005) Analysis of FOXP3 protein expression in human CD4+CD25+ regulatory T cells at the single-cell level. Eur J Immunol 35: 1681–1691.
90. Gavin MA, Torgerson TR, Houston E, DeRoos P, Ho WY, et al. (2006) Single-cell analysis of normal and FOXP3-mutant human T cells: FOXP3 expression without regulatory T cell development. Proc Natl Acad Sci U S A 103: 6659–6664.
91. Tran DQ, Ramsey H, Shevach EM (2007) Induction of FOXP3 expression in naive human CD4+FOXP3 T cells by T-cell receptor stimulation is transforming growth factor-beta dependent but does not confer a regulatory phenotype. Blood 110: 2983–2990.
92. Pillai V, Ortega SB, Wang CK, Karandikar NJ (2007) Transient regulatory T-cells: a state attained by all activated human T-cells. Clin Immunol 123: 18–29.
93. Wang J, Ioan-Facsinay A, van der Voort EI, Huizinga TW, Toes RE (2007) Transient expression of FOXP3 in human activated nonregulatory CD4+ T cells. Eur J Immunol 37: 129–138.
94. Allan SE, Crome SQ, Crellin NK, Passerini L, Steiner TS, et al. (2007) Activation-induced FOXP3 in human T effector cells does not suppress proliferation or cytokine production. International immunology 19: 345–354.
95. Imamichi H, Sereti I, Lane HC (2008) IL-15 acts as a potent inducer of CD4(+)CD25(hi) cells expressing FOXP3. Eur J Immunol 38: 1621–1630.
96. Roncarolo MG, Gregori S (2008) Is FOXP3 a bona fide marker for human regulatory T cells? Eur J Immunol 38: 925–927.
97. Kleinewietfeld M, Starke M, Di Mitri D, Borsellino G, Battistini L, et al. (2009) CD49d provides access to ''untouched'' human Foxp3+ Treg free of contaminating effector cells. Blood 113: 827–836.
98. Banham AH (2006) Cell-surface IL-7 receptor expression facilitates the purification of FOXP3(+) regulatory T cells. Trends Immunol 27: 541–544.
99. Alvaro T, Lejeune M, Salvado MT, Bosch R, Garcia JF, et al. (2005) Outcome in Hodgkin's lymphoma can be predicted from the presence of accompanying cytotoxic and regulatory T cells. Clin Cancer Res 11: 1467–1473.
100. Badoual C, Hans S, Rodriguez J, Peyrard S, Klein C, et al. (2006) Prognostic value of tumor-infiltrating CD4+ T-cell subpopulations in head and neck cancers. Clin Cancer Res 12: 465–472.
101. Lee AM, Clear AJ, Calaminici M, Davies AJ, Jordan S, et al. (2006) Number of CD4+ cells and location of forkhead box protein P3-positive cells in diagnostic follicular lymphoma tissue microarrays correlates with outcome. J Clin Oncol 24: 5052–5059.
102. Badoual C, Hans S, Fridman WH, Brasnu D, Erdman S, et al. (2009) Revisiting the Prognostic Value of Regulatory T Cells in Patients With Cancer. J Clin Oncol.
103. Hillen F, Baeten CI, van de Winkel A, Creytens D, van der Schaft DW, et al. (2008) Leukocyte infiltration and tumor cell plasticity are parameters of aggressiveness in primary cutaneous melanoma. Cancer Immunol Immunother 57: 97–106.
104. Ayyoub M, Deknuydt F, Raimbaud I, Dousset C, Leveque L, et al. (2009) Human memory FOXP3+ Tregs secrete IL-17 ex vivo and constitutively express the TH17 lineage-specific transcription factor ROR{gamma}t.
105. Voo KS, Wang YH, Santori FR, Boggiano C, Arima K, et al. (2009) Identification of IL-17-producing FOXP3+ regulatory T cells in humans. Proc Natl Acad Sci U S A 106: 4793–4798.
106. Koenen HJ, Smeets RL, Vink PM, van Rijssen E, Boots AM, et al. (2008) Human CD25highFoxp3pos regulatory T cells differentiate into IL-17-producing cells. Blood 112: 2340–2352.
107. Bettelli E, Carrier Y, Gao W, Korn T, Strom TB, et al. (2006) Reciprocal developmental pathways for the generation of pathogenic effector TH17 and regulatory T cells. Nature 441: 235–238.
108. Veldhoen M, Hocking RJ, Atkins CJ, Locksley RM, Stockinger B (2006) TGFbeta in the context of an inflammatory cytokine milieu supports de novo differentiation of IL-17-producing T cells. Immunity 24: 179–189.
109. Stockinger B, Veldhoen M (2007) Differentiation and function of Th17 T cells. Curr Opin Immunol 19: 281–286.
110. Kato T, Furumoto H, Ogura T, Onishi Y, Irahara M, et al. (2001) Expression of IL-17 mRNA in ovarian cancer. Biochem Biophys Res Commun 282: 735–738.
111. Miyahara Y, Odunsi K, Chen W, Peng G, Matsuzaki J, et al. (2008) Generation and regulation of human CD4+ IL-17-producing T cells in ovarian cancer. Proc Natl Acad Sci U S A 105: 15505–15510.
112. Leveque L, Deknuydt F, Bioley G, Old LJ, Matsuzaki J, et al. (2009) Interleukin 2-mediated conversion of ovarian cancer-associated CD4+ regulatory T cells into proinflammatory interleukin 17-producing helper T cells. J Immunother 32: 101–108.
113. Dong HP, Elstrand MB, Holth A, Silins I, Berner A, et al. (2006) NK- and B-cell infiltration correlates with worse outcome in metastatic ovarian carcinoma. Am J Clin Pathol 125: 451–458.
114. Hansen MH, Nielsen HV, Ditzel HJ (2002) Translocation of an intracellular antigen to the surface of medullary breast cancer cells early in apoptosis allows for an antigen-driven antibody response elicited by tumor-infiltrating B cells. J Immunol 169: 2701–2711.
115. Hansen MH, Nielsen H, Ditzel HJ (2001) The tumor-infiltrating B cell response in medullary breast cancer is oligoclonal and directed against the autoantigen actin exposed on the surface of apoptotic cancer cells. Proc Natl Acad Sci U S A 98: 12659–12664.
116. Schmidt M, Bohm D, von Torne C, Steiner E, Puhl A, et al. (2008) The humoral immune system has a key prognostic impact in node-negative breast cancer. Cancer Res 68: 5405–5413.
117. Wang Y, Ylera F, Boston M, Kang SG, Kutok JL, et al. (2007) Focused antibody response in plasma cell-infiltrated non-medullary (NOS) breast cancers. Breast Cancer Res Treat 104: 129–144.
118. Coronella JA, Spier C, Welch M, Trevor KT, Stopeck AT, et al. (2002) Antigen-driven oligoclonal expansion of tumor-infiltrating B cells in infiltrating ductal carcinoma of the breast. J Immunol 169: 1829–1836.
119. Kotlan B, Simsa P, Teillaud JL, Fridman WH, Toth J, et al. (2005) Novel ganglioside antigen identified by B cells in human medullary breast carcinomas: the proof of principle concerning the tumor-infiltrating B lymphocytes. J Immunol 175: 2278–2285.
120. Nzula S, Going JJ, Stott DI (2003) Antigen-driven clonal proliferation, somatic hypermutation, and selection of B lymphocytes infiltrating human ductal breast carcinomas. Cancer Res 63: 3275–3280.
121. Kurt-Jones EA, Liano D, HayGlass KA, Benacerraf B, Sy MS, et al. (1988) The role of antigen-presenting B cells in T cell priming in vivo. Studies of B cell-deficient mice. J Immunol 140: 3773–3778.
122. Yanaba K, Bouaziz JD, Matsushita T, Magro CM, St Clair EW, et al. (2008) B-lymphocyte contributions to human autoimmune disease. Immunol Rev 223: 284–299.
123. Constant S, Schweitzer N, West J, Ranney P, Bottomly K (1995) B lymphocytes can be competent antigen-presenting cells for priming CD4+ T cells to protein antigens in vivo. J Immunol 155: 3734–3741.
124. Ron Y, De Baetselier P, Gordon J, Feldman M, Segal S (1981) Defective induction of antigen-reactive proliferating T cells in B cell-deprived mice. Eur J Immunol 11: 964–968.

125. Ron Y, Sprent J (1987) T cell priming in vivo: a major role for B cells in presenting antigen to T cells in lymph nodes. J Immunol 138: 2848–2856.
126. Janeway CA Jr. Ron J, Katz ME (1987) The B cell is the initiating antigen-presenting cell in peripheral lymph nodes. J Immunol 138: 1051–1055.
127. Liu Y, Wu Y, Ramarathinam L, Guo Y, Huszar D, et al. (1995) Gene-targeted B-deficient mice reveal a critical role for B cells in the CD4 T cell response. Int Immunol 7: 1353–1362.
128. Linton PJ, Harbertson J, Bradley LM (2000) A critical role for B cells in the development of memory CD4 cells. J Immunol 165: 5558–5565.
129. Crawford A, Macleod M, Schumacher T, Corlett L, Gray D (2006) Primary T cell expansion and differentiation in vivo requires antigen presentation by B cells. J Immunol 176: 3498–3506.
130. Shen H, Whitmire JK, Fan X, Shedlock DJ, Kaech SM, et al. (2003) A specific role for B cells in the generation of CD8 T cell memory by recombinant Listeria monocytogenes. J Immunol 170: 1443–1451.
131. Ronchese F, Hausmann B (1993) B lymphocytes in vivo fail to prime naive T cells but can stimulate antigen-experienced T lymphocytes. J Exp Med 177: 679–690.
132. Tamiolakis D, Simopoulos C, Cheva A, Lambropoulou M, Kotini A, et al. (2002) Immunophenotypic profile of tumor infiltrating lymphocytes in medullary carcinoma of the breast. Eur J Gynaecol Oncol 23: 433–436.
133. Al-Shibli KI, Donnem T, Al-Saad S, Persson M, Bremnes RM, et al. (2008) Prognostic effect of epithelial and stromal lymphocyte infiltration in non-small cell lung cancer. Clin Cancer Res 14: 5220–5227.
134. Nedergaard BS, Ladekarl M, Nyengaard JR, Nielsen K (2008) A comparative study of the cellular immune response in patients with stage IB cervical squamous cell carcinoma. Low numbers of several immune cell subtypes are strongly associated with relapse of disease within 5 years. Gynecol Oncol 108: 106–111.
135. Bingle L, Brown NJ, Lewis CE (2002) The role of tumour-associated macrophages in tumour progression: implications for new anticancer therapies. J Pathol 196: 254–265.
136. Sica A, Larghi P, Mancino A, Rubino L, Porta C, et al. (2008) Macrophage polarization in tumour progression. Semin Cancer Biol 18: 349–355.
137. Gottfried E, Kunz-Schughart LA, Weber A, Rehli M, Peuker A, et al. (2008) Expression of CD68 in non-myeloid cell types. Scand J Immunol 67: 453–463.

Human Immunology 71 (2010) 245–254

Contents lists available at ScienceDirect

journal homepage: www.elsevier.com/locate/humimm

# A census of predicted mutational epitopes suitable for immunologic cancer control

René L. Warren [a] and Robert A. Holt [a,b,*]

[a] British Columbia Cancer Agency, Genome Sciences Centre, Vancouver, British Columbia, Canada
[b] Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, Canada

ARTICLE INFO

*Article history:*
Received 2 October 2009
Accepted 17 December 2009
Available online 9 January 2010

*Keywords:*
Epitope prediction
T-cell epitope
Immunosurveillance
Cancer vaccine
Cancer immunity

ABSTRACT

The adaptive immune system can protect against spontaneously arising tumors, and the potential exists to reduce cancer incidence by priming adaptive immune responses with vaccines. Immunologic cancer control has been implemented for cancers caused by infectious agents, but not for spontaneous cancers caused by mutation. This is largely due to the high cost of preventative clinical trials and the lack of validated tumor epitopes. Here we evaluate, computationally, all known somatic mutations in human tumors for their antigenic potential. All possible human leukocyte antigen (HLA) class I presented peptides containing recurrent somatic cancer mutations with frequency > 5% were screened by three independent epitope prediction algorithms (SYFPEITHI, BIMAS, and IEDB). Using stringent filters, a total of 20 genes, 35 mutations, and 159 candidate epitopes were identified, each presented by up to four distinct HLA class I alleles. The top-ranking gene from our survey was KRAS, which figures prominently because there are frequent hotspot mutations in numerous, prevalent cancers, and mutant peptides are predicted to be presented by several common HLA alleles. From our data, we estimate that prophylactic vaccination could provide meaningful levels of prevention of tumors associated with common recurrent mutations.



## 1. Introduction

More than a century ago it was recognized independently by Ehrlich as well as Bashford and colleagues [1,2] that cells isolated from rodent tumors could be lethal when injected into naive mice but were unable to proliferate when injected at a new site in the same animal from which they were obtained, even as the original tumor continued to grow. This phenomenon was termed concomitant immunity and clearly illustrated that the mammalian immune system is effective in eliminating cancer if the burden of malignant cells is low, but is less effective for established tumors. Half a century later, a series of definitive experiments by Prehn and Main [3] showed that inoculation with tumor, but not normal tissue, was protective and, in time, the theory of cancer immunosurveillance was formally established [4]. Recently, it has been shown that Rag2$^{-/-}$ mice, which have no V(D)J recombination and therefore no repertoire of mature lymphocytes, have dramatically increased incidence of spontaneous tumors [5]. Further, it has been shown that tumors formed in immunologically permissive Rag2$^{-/-}$ hosts are more immunogenic when transferred to wild-type mice, and that host-immune pressures can maintain tumors in an equilibrium state [6]. Based on these observations, the early theory of immunosurveillance has been revised to that of immunoediting [7], which holds that spontaneously arising tumor cells are frequently eliminated by the immune system. Those that begin to grow are held in a state equilibrium with host immunity, from which they may eventually escape by various mechanisms such as loss of antigen, loss of antigen presentation pathways, or interference by regulatory T cells. The potential to use vaccination to mobilize adaptive immunity against cancer has been illustrated in mice engineered to express, for example, the SV40 T antigen [8], or activated rat Erbb2 [9–12]. In these animal models, pre-exposure to antigen can reproducibly provide complete protection against tumor development.

In humans, it is already well established that cancers of viral origin can be prevented by vaccination, and Gardasil® (Merck & Co., Inc., Whitehouse Station, NJ) now makes this a clinical reality for cervical cancer [13]. For tumors of mutational origin, the principal lines of evidence for antitumor adaptive immune responses are as follows. First, there are numerous case reports of donated organs in which occult cancers, initially held in check by donor immunity, undergo rapid and progressive outgrowth after transplantation because recipients are naive to the tumor antigens, and immunosuppressed [14–16]. Second, the only controlled study that has evaluated cancer rates in immunosuppressed individuals (Scandinavian kidney transplant recipients) reports a clear increase in the incidence of a wide variety of noninfectious primary cancers [17]. Third, it is well established in solid tumors that patients with detectable tumor infiltrating lymphocytes have better outcomes,

* Corresponding author.
E-mail address: rholt@bcgsc.ca (R.A. Holt).

0198-8859/10/$32.00 - see front matter © 2010 American Society for Histocompatibility and Immunogenetics. Published by Elsevier Inc. All rights reserved.
doi:10.1016/j.humimm.2009.12.007

both in terms of progression free interval and overall survival [18–20]. Finally, natural immune responses to tumor-specific antigens are detectable in patients with cancer. Humoral responses have been extensively characterized in patient serum and antibodies against common tumor antigens have been found, in some studies, in more than half of cancer patients but in very few healthy control individuals (reviewed in [21]). Natural cellular immune responses are more difficult to assess, and efforts have focused on characterization of tumor-infiltrating rather than circulating lymphocytes. $CD4^+$ and $CD8^+$ tumor-reactive T cells have been identified that recognize point mutations in numerous genes such as fibronectin [22], HSP70 [23], major histocompatibility complex class I [24], N-RAS [25], $\beta$-catenin [26], MART-2 [27], and p14ARF [24].

It is conspicuous that after more than a century of studying cancer immunology we still do not have immunologic cancer control (*i.e.*, vaccines) [28] for the most common and deadly cancers. The reasons for this include the fact that it is difficult to test preventative strategies against cancer. The likelihood of any individual developing a specific type of tumor is relatively low, and when they do appear, they occur over a wide age range such that preventative clinical trials require the involvement of a large number of subjects over many years. These trials are, therefore, expensive and require extraordinary justification. Further, for most cancers we do not have validated antigens. In cancers of viral origin, the human immune system responds to non-self antigens from the virus. In contrast, vaccination against tumors of mutational origin involves important and unique considerations. Antigens must be sufficiently common in the target cancer to have a meaningful public health impact and they must be reliably immunogenic (*i.e.*, nontolerated). They should also be causally involved in tumorigenicity and confer a selective advantage necessary for tumor survival, such that the risk of immune escape by antigen loss is minimized.

Most oncogenic mutations are incurred in intracellular signaling proteins. In nucleated cells, intracellular proteins are cleaved and a subset of the cleavage products are presented at the cell surface by human leukocyte antigen (HLA) class I, where they are subjected to surveillance by the repertoire of cytotoxic T lymphocytes. Because HLA class I restricted peptides are short (~8–11 amino acids), even single amino acid changes such as the hotspot mutations incurred by common tumor genes can be sufficient for cytotoxic T lymphocyte to single out and selectively kill cells presenting the mutant peptide [29,30]. For these reasons mutational epitopes (rather than expression or differentiation epitopes) will likely prove to be best suited for preventative vaccines against spontaneous cancers. As of this writing, only 39 HLA class I restricted mutational epitopes have been reported and compiled by the cancer immunity database (http://www.cancerimmunity.org). In this set there is over-representation of melanoma epitopes (17/39) and epitopes that are HLA-A2 restricted (11/39). HLA loci, particularly HLA antigen-binding regions, are the most highly polymorphic sites in the genome with 3,477 HLA class I allelic sequences now recognized (http://hla.alleles.org/alleles/index.html). Different HLA alleles have different antigen presentation characteristics, such that presentation of any given peptide is often restricted to, or strongly favored by, a specific HLA allele [31]. HLA restriction is a key consideration when assessing the immunogenicity of oncogenic mutations.

Here, we present a meta-analysis of predicted HLA class I restricted epitopes from recurrent tumor mutations using established databases and epitope prediction tools. We have evaluated, computationally, all mutant peptides derived from all known tumor mutations. Epitopes have been ranked according to (1) the estimated frequency of mutation in a given tumor type, (2) the incidence of that tumor, (3) the likelihood of the mutation producing an HLA presented T-cell epitope, and (4) the estimated population frequency of the presenting HLA allele. Because of the stringency of our analysis our predictions are very conservative, but offer the first quantitative estimate of the minimum benefit that might be expected from preventative cancer vaccination, and provide a short list of the best predicted epitopes for further study.

## 2. Methods

### 2.1. Selection of candidate tumor mutations

Researchers at the Sanger Institute [32–34] have undertaken the daunting task of curating and cataloging over 50,000 published somatic mutations from almost 4,800 genes in 250,000 tumors scattered across the vast cancer literature. Their database, COSMIC (Catalogue of Somatic Mutations in Cancer), tracks each gene and somatic change observed, its frequency, and the primary tissue where it occurs. Using COSMIC v43 Release (http://www.sanger.ac.uk/genetics/CGP/cosmic/), every combination of primary tissue, gene, and somatic mutation was considered for our study. Only somatic mutations seen in 5% or more of tissue samples and where at least five or more positives identified from 100 or more samples were retained for further analysis. A further condition for tumor mutation selection was that the cancer incidence be known for each tissue type where the mutation occurred. Cancer incidence data from the Surveillance, Epidemiology and End Results (SEER) program (http://seer.cancer.gov), which is the definitive source for cancer statistics in the United States, was used for this purpose.

First, we defined a Mutation Impact Score (MIS) as follows:

$$\text{MIS} = \text{Frequency of mutation in given tumor} \times \text{tumor incidence}$$

Each mutation was considered independently, regardless of whether there were other mutations in the same gene, or even at the same codon (*e.g.*, KRAS G12V and KRAS G12D were treated as independent mutations). For comparing mutations we also calculated a cumulative MIS, which was the sum of the MIS for all tumor sites in which that mutation was detected (Figure 1).

### 2.2. Identification of putative HLA ligands

We used computational HLA-binding prediction tools to identify mutations likely to be presented by specific HLA alleles as peptide ligands and, by extension, T-cell epitopes. Using web-based SYFPEITHI [35] (http://www.syfpeithi.de), BIMAS [36] (http://www-bimas.cit.nih.gov/molbio/hla_bind/), and IEDB [37,38] (Stabilized Matrix Method; http://tools.immuneepitope.org/analyze/html/mhc_binding.html), we queried each candidate mutation in a systematic and high-throughput fashion, testing all combinations of peptide lengths and HLA class I alleles available to query on the hosting server. These prediction algorithms use empiric peptide-HLA binding data to build models of peptide sequence specificity, and give a predictive outcome for yet untested sequences. In these models, the amino acids found at each position of short peptides eluted from real HLA molecules are compiled in a database. For SYFPEITHI and BIMAS a two-dimensional matrix is built using the observed frequency of each amino acid at each position within the HLA peptide binding pocket. Stabilized matrix method (IEDB) considers, in addition, pairwise interactions between amino acid positions and is thought to yield more accurate predictions [38]. These three tools are the most heavily used and highly cited tools for *in silico* HLA-binding predictions [39].

A script using PERL LWP and HTML::Form modules was developed to handle high-throughput form submission and post-prediction data extraction. This script was run in a manner that minimized impact on the hosting service, with a 5-second buffer between each submission. All overlapping 8-, 9-, 10-, and 11-mer mutational peptides were tested against all HLA class I combinations supported by the host servers. For each peptide containing the mutation of interest, the output in terms of HLA-binding prediction method, rank, score,

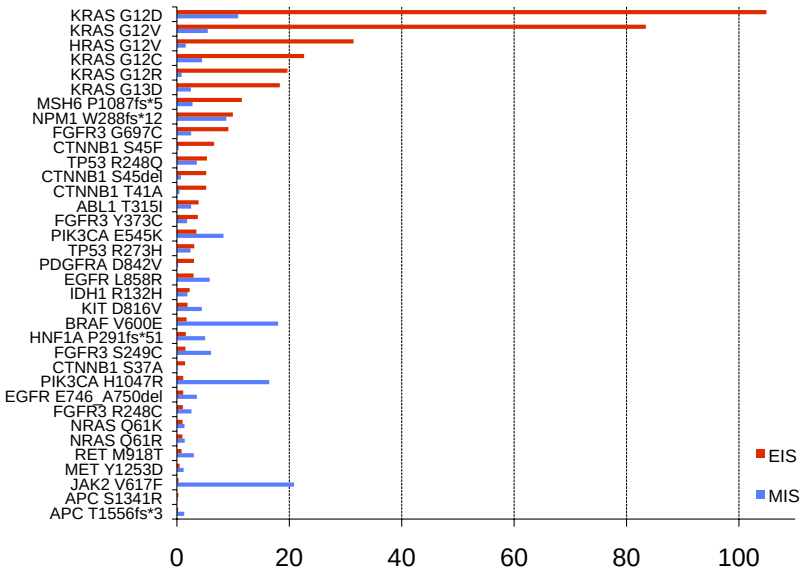



**Fig. 1.** Sum of mutation impact score (MIS) and sum of epitope impact score (EIS) for each candidate tumor-associated antigen identified by our study. The EIS, which takes into account the average HLA frequency in the population, and the HLA-binding prediction ranking is high for all KRAS mutants, which suggests that their predicted epitopes would be expected to have the greatest utility for immunologic cancer control.

start coordinate, length, peptide sequence, and HLA allele were stored in a local custom MySQL database for future querying. To be considered further, each peptide-HLA allele combination had to be found by at least two of the three epitope prediction algorithms and the predicted HLA ligand had to be within the top scoring 5% of all possible peptides from that protein. Only HLA class I coding variants were considered (*i.e.*, 4-digit resolution). Candidate epitopes that passed this filter were given an epitope score (ES).

$$\mathrm{ES} = 100/(\mathrm{AVG\ RANK} \times \mathrm{EXP}(\mathrm{STDEV}/100))$$

The ES is based on the mean prediction rank and standard deviation from the mean for two or more prediction algorithms and ranges from 0 to 100. A peptide predicted to bind to a given HLA and ranked first by any two or all three prediction programs will score 100 (*e.g.*, KRAS G12R VVVGARGVGK). We used the rank as a metric because the epitope prediction scores are determined differently by the three programs and are not directly comparable.

Next, we calculated Epitope Impact Score (EIS).

$$\mathrm{EIS} = \mathrm{ES} \times \text{average population frequency of presenting HLA allele}$$

The EIS considers the US population frequencies for each HLA class I allele (http://www.allelefrequencies.net). Only allelic frequencies with a sample size equal to or greater than 500 were used. The EIS will decrease the ES by a factor proportional to the allele frequency itself. Thus, even high-ranking peptides will have a low EIS score if they are predicted to be presented by rare HLA alleles. For example, top-ranking KRAS G12R peptide VVVGARGVGK (ES = 100), predicted to be presented by HLA class I coding variant A*1101 (average population frequency = 7.07%) has a low EIS of 7.07.

For each candidate peptide-HLA pair we calculated a global score (GS) that takes into account the MIS as

$$\mathrm{GS} = \mathrm{MIS} \times \mathrm{EIS}$$

And for each gene, we calculated an overall score (OS) which is the sum of the GS for all predicted epitopes for that gene across all tumor sites where the mutation is found.

$$\mathrm{OS} = \sum \mathrm{GS}$$

To estimate the proportion of the population that could benefit from vaccination with a combination of predicted epitopes, it is necessary to determine the proportion of the population that would carry at least one presenting HLA allele. The proportion (P) of the population that contains at least one of some larger number of alleles (K) is determined as described in Gulukota and DeLisi [40]. Assuming that two or more alleles occur with correlated probabilities, the overall coverage is the sum of individual allele coverage corrected for the overlaps.

$$P_{tot} = \sum_{i=1}^{i=k} P_i - \sum_{pairs} P_{ij} + \sum_{triplets} P_{ijk} - \ldots,$$

## 3. Results

We used the COSMIC database as a starting point to identify the most promising tumor mutations for epitope prediction. COSMIC compiles data on somatic tumor mutations from the public scientific literature, and it is the most comprehensive repository of curated somatic mutations in cancer [32–34]. Using our selection criteria for somatic mutations (minimum frequency 5%, minimum of five positives from at least 100 samples) we identified 36 distinct somatic mutations in 20 genes and 19 distinct tumor sites (Table 1). These consisted of point mutations (29 cases), indels causing frameshifts (five cases) and codon deletions (two cases). These results demonstrate that based on mutational screens undertaken to date, there appear to be relatively few recurrent mutations in cancer. The highest scoring mutation was JAK2 V617F in hematopoietic/lymphoid tumors (MIS = 20.8), followed by PIK3CA H1047R in breast tumors (MIS = 15.1), and the NPM1 W288-frameshift in, again, hematopoietic/lymphoid tumors (MIS = 8.8). Interestingly, JAK2 is prominent because it is mutated at high frequency (45.1%) in a single category of common tumors (hematopoietic and lymphoid). NPM1, though also mutated only in this same category of tumors, ranks lower because of lower mutation frequency (19%).

Our screen prioritizes genes where mutations occur at specific locations and in multiple cancers. These are the candidates where the smallest number of epitopes could have the greatest potential impact when delivered as a vaccine. As a result, there are some well-recognized cancer genes that do not figure prominently in our results. For example, TP53 is one of the most frequently mutated cancer genes, but it does not rank highly here because mutations occur at many sites throughout the gene rather than at one or a small number of specific hotspots. For the present study a low TP53

**Table 1**
Predicted mutational tumor epitopes from cancer genes sorted by overall score for each gene. The mutated amino acid is underlined

| Gene | Mutation[a] | Tissue | Cancer incidence[b] (per 100k) | Mutation frequency[c] | MIS[d] | Peptide | Presenting allele | HLA frequency[e] | EIS[f] | GS[g] | OS[h] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| KRAS | G13D | Large intestine | 45.94 | 5.42% | 2.49 | VVVGAGDVGK | A*1101 | 7.07% | 5.272 | 13.127 | 506.52 |
| | | | | | | | A*6801 | 3.50% | 0.671 | 1.670 | |
| | | | | | | KLVVVGAGDV | A*0201 | 13.66% | 5.349 | 13.319 | |
| | | | | | | GAGDVGKSAL | B*3801 | 1.03% | 0.221 | 0.549 | |
| | | | | | | VVGAGDVGK | A*1101 | 7.07% | 2.089 | 5.201 | |
| | | | | | | AGDVGKSAL | B*0702 | 6.71% | 3.307 | 8.235 | |
| | | | | | | | B*3801 | 1.03% | 0.227 | 0.565 | |
| | | | | | | | B*3901 | 0.87% | 0.120 | 0.299 | |
| | | | | | | DVGKSALTI | B*5101 | 4.29% | 1.057 | 2.631 | |
| | G12V | Biliary tract | 1.14 | 5.89% | 0.07 | VVGAVGVGK | A*1101 | 7.07% | 3.013 | 0.202 | |
| | | | | | | GAVGVGKSAL | B*0702 | 6.71% | 1.072 | 0.072 | |
| | | | | | | | B*3801 | 1.03% | 0.221 | 0.015 | |
| | | | | | | KLVVVGAVGV | A*0201 | 13.66% | 8.102 | 0.544 | |
| | | | | | | AVGVGKSAL | B*0702 | 6.71% | 4.441 | 0.298 | |
| | | | | | | VVVGAVGVGK | A*1101 | 7.07% | 7.070 | 0.475 | |
| | | | | | | | A*6801 | 3.50% | 0.751 | 0.050 | |
| | | | | | | YKLVVVGAV | A*0203 | 0.90% | 0.256 | 0.017 | |
| | | | | | | | B*3902 | 0.11% | 0.017 | 0.001 | |
| | | | | | | LVVVGAVGV | A*0201 | 13.66% | 2.732 | 0.183 | |
| | | | | | | | A*0203 | 0.90% | 0.138 | 0.009 | |
| | | Large intestine | 45.94 | 7.29% | 3.35 | VVGAVGVGK | A*1101 | 7.07% | 3.013 | 10.089 | |
| | | | | | | GAVGVGKSAL | B*0702 | 6.71% | 1.072 | 3.589 | |
| | | | | | | | B*3801 | 1.03% | 0.221 | 0.738 | |
| | | | | | | KLVVVGAVGV | A*0201 | 13.66% | 8.102 | 27.133 | |
| | | | | | | AVGVGKSAL | B*0702 | 6.71% | 4.441 | 14.873 | |
| | | | | | | VVVGAVGVGK | A*1101 | 7.07% | 7.070 | 23.678 | |
| | | | | | | | A*6801 | 3.50% | 0.751 | 2.514 | |
| | | | | | | YKLVVVGAV | A*0203 | 0.90% | 0.256 | 0.859 | |
| | | | | | | | B*3902 | 0.11% | 0.017 | 0.056 | |
| | | | | | | LVVVGAVGV | A*0201 | 13.66% | 2.732 | 9.149 | |
| | | | | | | | A*0203 | 0.90% | 0.138 | 0.462 | |
| | | Pancreas | 11.70 | 17.67% | 2.07 | VVGAVGVGK | A*1101 | 7.07% | 3.013 | 6.228 | |
| | | | | | | GAVGVGKSAL | B*0702 | 6.71% | 1.072 | 2.216 | |
| | | | | | | | B*3801 | 1.03% | 0.221 | 0.456 | |
| | | | | | | KLVVVGAVGV | A*0201 | 13.66% | 8.102 | 16.749 | |
| | | | | | | AVGVGKSAL | B*0702 | 6.71% | 4.441 | 9.182 | |
| | | | | | | VVVGAVGVGK | A*1101 | 7.07% | 7.070 | 14.616 | |
| | | | | | | | A*6801 | 3.50% | 0.751 | 1.552 | |
| | | | | | | YKLVVVGAV | A*0203 | 0.90% | 0.256 | 0.530 | |
| | | | | | | | B*3902 | 0.11% | 0.017 | 0.035 | |
| | | | | | | LVVVGAVGV | A*0201 | 13.66% | 2.732 | 5.648 | |
| | | | | | | | A*0203 | 0.90% | 0.138 | 0.285 | |
| | G12R | Pancreas | 11.70 | 7.17% | 0.84 | ARGVGKSAL | B*0702 | 6.71% | 1.608 | 1.349 | |
| | | | | | | | B*2705 | 1.38% | 0.331 | 0.278 | |
| | | | | | | | B*3901 | 0.87% | 0.239 | 0.200 | |
| | | | | | | GARGVGKSAL | B*0702 | 6.71% | 1.286 | 1.079 | |
| | | | | | | EYKLVVVGAR | A*3101 | 2.40% | 0.600 | 0.503 | |
| | | | | | | KLVVVGARGV | A*0201 | 13.66% | 3.633 | 3.048 | |
| | | | | | | VVGARGVGK | A*1101 | 7.07% | 3.013 | 2.527 | |
| | | | | | | RGVGKSALTI | B*0702 | 6.71% | 1.087 | 0.912 | |
| | | | | | | VVVGARGVGK | A*1101 | 7.07% | 7.070 | 5.931 | |
| | | | | | | | A*6801 | 3.50% | 0.671 | 0.563 | |
| | | | | | | LVVVGARGV | A*0203 | 0.90% | 0.138 | 0.116 | |
| | G12D | Biliary tract | 1.14 | 15.43% | 0.18 | VVGADGVGK | A*1101 | 7.07% | 2.333 | 0.410 | |
| | | | | | | LVVVGADGV | A*0201 | 13.66% | 1.683 | 0.296 | |
| | | | | | | | A*0203 | 0.90% | 0.127 | 0.022 | |
| | | | | | | KLVVVGADGV | A*0201 | 13.66% | 6.713 | 1.181 | |
| | | | | | | GADGVGKSAL | B*3801 | 1.03% | 0.680 | 0.120 | |
| | | | | | | VVVGADGVGK | A*1101 | 7.07% | 5.272 | 0.927 | |
| | | | | | | | A*6801 | 3.50% | 0.671 | 0.118 | |
| | | Endometrium | 23.49 | 5.67% | 1.33 | VVGADGVGK | A*1101 | 7.07% | 2.333 | 3.108 | |
| | | | | | | LVVVGADGV | A*0201 | 13.66% | 1.683 | 2.242 | |
| | | | | | | | A*0203 | 0.90% | 0.127 | 0.170 | |
| | | | | | | KLVVVGADGV | A*0201 | 13.66% | 6.713 | 8.940 | |
| | | | | | | GADGVGKSAL | B*3801 | 1.03% | 0.680 | 0.906 | |
| | | | | | | VVVGADGVGK | A*1101 | 7.07% | 5.272 | 7.022 | |
| | | | | | | | A*6801 | 3.50% | 0.671 | 0.893 | |
| | | Large intestine | 45.94 | 11.32% | 5.20 | VVGADGVGK | A*1101 | 7.07% | 2.333 | 12.134 | |
| | | | | | | LVVVGADGV | A*0201 | 13.66% | 1.683 | 8.755 | |
| | | | | | | | A*0203 | 0.90% | 0.127 | 0.662 | |
| | | | | | | KLVVVGADGV | A*0201 | 13.66% | 6.713 | 34.908 | |
| | | | | | | GADGVGKSAL | B*3801 | 1.03% | 0.680 | 3.539 | |
| | | | | | | VVVGADGVGK | A*1101 | 7.07% | 5.272 | 27.416 | |
| | | | | | | | A*6801 | 3.50% | 0.671 | 3.488 | |

**Table 1**
(*continued*)

| Gene | Mutation[a] | Tissue | Cancer incidence[b] (per 100k) | Mutation frequency[c] | MIS[d] | Peptide | Presenting allele | HLA frequency[e] | EIS[f] | GS[g] | OS[h] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Ovary | 12.62 | 5.47% | 0.69 | VVGADGVGK | A*1101 | 7.07% | 2.333 | 1.611 | |
| | | | | | | LVVVGADGV | A*0201 | 13.66% | 1.683 | 1.162 | |
| | | | | | | | A*0203 | 0.90% | 0.127 | 0.088 | |
| | | | | | | KLVVVGADGV | A*0201 | 13.66% | 6.713 | 4.634 | |
| | | | | | | GADGVGKSAL | B*3801 | 1.03% | 0.680 | 0.470 | |
| | | | | | | VVVGADGVGK | A*1101 | 7.07% | 5.272 | 3.639 | |
| | | | | | | | A*6801 | 3.50% | 0.671 | 0.463 | |
| | | Pancreas | 11.70 | 28.57% | 3.34 | VVGADGVGK | A*1101 | 7.07% | 2.333 | 7.799 | |
| | | | | | | LVVVGADGV | A*0201 | 13.66% | 1.683 | 5.627 | |
| | | | | | | | A*0203 | 0.90% | 0.127 | 0.426 | |
| | | | | | | KLVVVGADGV | A*0201 | 13.66% | 6.713 | 22.438 | |
| | | | | | | GADGVGKSAL | B*3801 | 1.03% | 0.680 | 2.275 | |
| | | | | | | VVVGADGVGK | A*1101 | 7.07% | 5.272 | 17.622 | |
| | | | | | | | A*6801 | 3.50% | 0.671 | 2.242 | |
| | | Small intestine | 1.92 | 8.33% | 0.16 | VVGADGVGK | A*1101 | 7.07% | 2.333 | 0.373 | |
| | | | | | | LVVVGADGV | A*0201 | 13.66% | 1.683 | 0.269 | |
| | | | | | | | A*0203 | 0.90% | 0.127 | 0.020 | |
| | | | | | | KLVVVGADGV | A*0201 | 13.66% | 6.713 | 1.074 | |
| | | | | | | GADGVGKSAL | B*3801 | 1.03% | 0.680 | 0.109 | |
| | | | | | | VVVGADGVGK | A*1101 | 7.07% | 5.272 | 0.843 | |
| | | | | | | | A*6801 | 3.50% | 0.671 | 0.107 | |
| | G12C | Lung | 60.66 | 7.38% | 4.48 | ACGVGKSAL | B*0702 | 6.71% | 1.850 | 8.283 | |
| | | | | | | VVVGACGVGK | A*1101 | 7.07% | 7.070 | 31.650 | |
| | | | | | | | A*6801 | 3.50% | 0.671 | 3.003 | |
| | | | | | | LVVVGACGV | A*0201 | 13.66% | 2.245 | 10.049 | |
| | | | | | | | A*0203 | 0.90% | 0.151 | 0.674 | |
| | | | | | | VVGACGVGK | A*1101 | 7.07% | 2.333 | 10.445 | |
| | | | | | | GACGVGKSAL | B*3801 | 1.03% | 0.221 | 0.987 | |
| | | | | | | KLVVVGACGV | A*0201 | 13.66% | 8.102 | 36.269 | |
| NPM1 | W288fs*12 | Hematopoietic and lymphoid tissue | 46.19 | 19.03% | 8.79 | DLCLAVEEV | A*0201 | 13.66% | 1.635 | 14.368 | 87.49 |
| | | | | | | DQEAIQDLCL | B*3801 | 1.03% | 0.118 | 1.041 | |
| | | | | | | EAIQDLCLA | A*6801 | 3.50% | 0.250 | 2.197 | |
| | | | | | | | B*4501 | 2.35% | 0.261 | 2.295 | |
| | | | | | | AIQDLCLAV | A*0201 | 13.66% | 4.051 | 35.607 | |
| | | | | | | QEAIQDLCL | B*4001 | 3.26% | 0.924 | 8.121 | |
| | | | | | | | B*4403 | 5.12% | 0.711 | 6.247 | |
| | | | | | | QEAIQDLCLA | B*4403 | 5.12% | 2.004 | 17.615 | |
| HRAS | G12V | Urinary tract | 20.27 | 7.73% | 1.57 | VVGAVGVGK | A*1101 | 7.07% | 4.218 | 6.608 | 49.24 |
| | | | | | | GAVGVGKSAL | B*0702 | 6.71% | 1.102 | 1.727 | |
| | | | | | | | B*3801 | 1.03% | 0.250 | 0.391 | |
| | | | | | | KLVVVGAVGV | A*0201 | 13.66% | 8.102 | 12.694 | |
| | | | | | | AVGVGKSAL | B*0702 | 6.71% | 6.709 | 10.512 | |
| | | | | | | VVVGAVGVGK | A*1101 | 7.07% | 7.070 | 11.078 | |
| | | | | | | | A*6801 | 3.50% | 0.838 | 1.314 | |
| | | | | | | YKLVVVGAV | A*0203 | 0.90% | 0.256 | 0.402 | |
| | | | | | | LVVVGAVGV | A*0201 | 13.66% | 2.732 | 4.281 | |
| | | | | | | | A*0203 | 0.90% | 0.151 | 0.236 | |
| FGFR3 | Y373C | Urinary tract | 20.27 | 8.95% | 1.81 | DEAGSVCAG | B*4403 | 5.12% | 0.188 | 0.341 | 42.06 |
| | | | | | | DEAGSVCAGI | B*4402 | 2.85% | 0.214 | 0.388 | |
| | | | | | | | B*4403 | 5.12% | 0.557 | 1.010 | |
| | | | | | | SVCAGILSY | A*1101 | 7.07% | 0.307 | 0.557 | |
| | | | | | | | B*1501 | 2.64% | 1.749 | 3.173 | |
| | | | | | | CAGILSYGV | B*5101 | 4.29% | 0.144 | 0.261 | |
| | | | | | | GSVCAGILSY | A*1101 | 7.07% | 0.264 | 0.479 | |
| | | | | | | EAGSVCAGI | B*5101 | 4.29% | 0.300 | 0.544 | |
| | S249C | Urinary tract | 20.27 | 29.97% | 6.07 | ERCPHRPIL | B*2705 | 1.38% | 0.064 | 0.389 | |
| | | | | | | | B*3901 | 0.87% | 0.054 | 0.326 | |
| | | | | | | LERCPHRPI | B*4403 | 5.12% | 0.136 | 0.828 | |
| | | | | | | TYTLDVLERC | A*2402 | 9.29% | 0.247 | 1.502 | |
| | | | | | | DVLERCPHR | A*1101 | 7.07% | 0.224 | 1.359 | |
| | | | | | | | A*3101 | 2.40% | 0.115 | 0.702 | |
| | | | | | | | A*6801 | 3.50% | 0.376 | 2.281 | |
| | | | | | | CPHRPILQA | B*0702 | 6.71% | 0.311 | 1.889 | |
| | R248C | Skin | 19.79 | 13.07% | 2.59 | TYTLDVLECS | A*2402 | 9.29% | 0.247 | 0.639 | |
| | | | | | | DVLECSPHR | A*1101 | 7.07% | 0.224 | 0.579 | |
| | | | | | | | A*3101 | 2.40% | 0.095 | 0.245 | |
| | | | | | | | A*6801 | 3.50% | 0.312 | 0.808 | |
| | | | | | | LECSPHRPI | B*4403 | 5.12% | 0.190 | 0.492 | |
| | G697C | Upper aerodigestive tract | 13.51 | 18.80% | 2.54 | CIPVEELFK | A*1101 | 7.07% | 0.408 | 1.035 | |
| | | | | | | LGGSPYPCI | B*5101 | 4.29% | 0.163 | 0.413 | |
| | | | | | | YPCIPVEELF | A*2402 | 9.29% | 0.266 | 0.674 | |
| | | | | | | | B*3501 | 5.82% | 0.916 | 2.327 | |
| | | | | | | PCIPVEELF | A*2402 | 9.29% | 0.492 | 1.250 | |
| | | | | | | TLGGSPYPCI | A*0201 | 13.66% | 0.455 | 1.156 | |

**Table 1**
(*continued*)

| Gene | Mutation[a] | Tissue | Cancer incidence[b] (per 100k) | Mutation frequency[c] | MIS[d] | Peptide | Presenting allele | HLA frequency[e] | EIS[f] | GS[g] | OS[h] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | CIPVEELFKL | A*0201 | 13.66% | 0.963 | 2.446 | |
| | | | | | | | B*3801 | 1.03% | 0.030 | 0.075 | |
| | | | | | | PYPCIPVEE | A*2402 | 9.29% | 0.430 | 1.092 | |
| | | | | | | PYPCIPVEEL | A*2402 | 9.29% | 3.637 | 9.238 | |
| | | | | | | YPCIPVEEL | B*0702 | 6.71% | 0.793 | 2.013 | |
| | | | | | | | B*3501 | 5.82% | 0.404 | 1.026 | |
| | | | | | | | B*5101 | 4.29% | 0.206 | 0.523 | |
| MSH6 | P1087fs*5 | Large intestine | 45.94 | 6.07% | 2.79 | LPEDTPPLL | B*0702 | 6.71% | 0.876 | 2.442 | 32.21 |
| | | | | | | | B*3501 | 5.82% | 0.260 | 0.724 | |
| | | | | | | | B*3801 | 1.03% | 0.053 | 0.148 | |
| | | | | | | | B*5101 | 4.29% | 0.443 | 1.236 | |
| | | | | | | ILLPEDTPPL | A*0201 | 13.66% | 4.475 | 12.479 | |
| | | | | | | LLPEDTPPLL | A*0201 | 13.66% | 1.631 | 4.548 | |
| | | | | | | | A*2402 | 9.29% | 0.878 | 2.449 | |
| | | | | | | | B*3801 | 1.03% | 0.026 | 0.071 | |
| | | | | | | LLPEDTPPL | A*0201 | 13.66% | 2.910 | 8.115 | |
| TP53 | R273H | Large intestine | 45.94 | 5.31% | 2.44 | GRNSFEVHV | B*2705 | 1.38% | 0.108 | 0.264 | 26.52 |
| | | | | | | HVCACPGRDR | A*3101 | 2.40% | 0.209 | 0.510 | |
| | | | | | | | A*6801 | 3.50% | 0.322 | 0.785 | |
| | | | | | | EVHVCACPGR | A*1101 | 7.07% | 0.610 | 1.489 | |
| | | | | | | | A*6801 | 3.50% | 1.725 | 4.208 | |
| | | | | | | GRNSFEVHVC | B*2705 | 1.38% | 0.119 | 0.291 | |
| | R248Q | Hematopoietic and lymphoid tissue | 46.19 | 7.69% | 3.55 | QRPILTIITL | A*2402 | 9.29% | 0.587 | 2.084 | |
| | | | | | | MGGMNQRPIL | B*5101 | 4.29% | 0.352 | 1.251 | |
| | | | | | | SCMGGMNQR | A*3101 | 2.40% | 0.269 | 0.954 | |
| | | | | | | MGGMNQRPI | B*5101 | 4.29% | 0.778 | 2.765 | |
| | | | | | | SSCMGGMNQR | A*1101 | 7.07% | 0.859 | 3.052 | |
| | | | | | | | A*6801 | 3.50% | 0.301 | 1.068 | |
| | | | | | | GMNQRPILTI | A*0201 | 13.66% | 2.196 | 7.800 | |
| EGFR | L858R | Lung | 60.66 | 9.60% | 5.82 | ITDFGRAKLL | B*3801 | 1.03% | 0.054 | 0.312 | 21.28 |
| | | | | | | HVKITDFGR | A*1101 | 7.07% | 0.252 | 1.470 | |
| | | | | | | | A*3101 | 2.40% | 1.200 | 6.988 | |
| | | | | | | | A*6801 | 3.50% | 0.084 | 0.491 | |
| | | | | | | KITDFGRAK | A*1101 | 7.07% | 0.249 | 1.449 | |
| | | | | | | RAKLLGAEEK | A*1101 | 7.07% | 0.447 | 2.600 | |
| | | | | | | KITDFGRAKL | A*0201 | 13.66% | 0.594 | 3.461 | |
| | | | | | | GRAKLLGAE | B*2705 | 1.38% | 0.026 | 0.153 | |
| | | | | | | ITDFGRAKL | B*3801 | 1.03% | 0.062 | 0.359 | |
| | E746_A750del | Lung | 60.66 | 5.86% | 3.55 | IPVAIKTSPK | A*1101 | 7.07% | 0.240 | 0.854 | |
| | | | | | | PVAIKTSPK | A*1101 | 7.07% | 0.287 | 1.020 | |
| | | | | | | AIKTSPKANK | A*1101 | 7.07% | 0.418 | 1.485 | |
| | | | | | | TSPKANKEI | A*2402 | 9.29% | 0.180 | 0.641 | |
| PIK3CA | H1047R | Breast | 120.81 | 12.51% | 15.11 | FMKQMNDAR | A*3101 | 2.40% | 0.085 | 1.279 | 18.84 |
| | | | | | | YFMKQMNDAR | A*3101 | 2.40% | 0.106 | 1.603 | |
| | | | | | | ARHGGWTTK | B*2705 | 1.38% | 0.336 | 5.073 | |
| | | | | | | ARHGGWTTKM | B*2705 | 1.38% | 0.037 | 0.557 | |
| | | Endometrium | 23.49 | 5.59% | 1.31 | FMKQMNDAR | A*3101 | 2.40% | 0.085 | 0.111 | |
| | | | | | | YFMKQMNDAR | A*3101 | 2.40% | 0.106 | 0.139 | |
| | | | | | | ARHGGWTTK | B*2705 | 1.38% | 0.336 | 0.441 | |
| | | | | | | ARHGGWTTKM | B*2705 | 1.38% | 0.037 | 0.048 | |
| | E545K | Breast | 120.81 | 5.39% | 6.51 | SEITKQEKD | B*4403 | 5.12% | 0.112 | 0.731 | |
| | | | | | | SEITKQEKDF | B*4402 | 2.85% | 0.374 | 2.436 | |
| | | | | | | | B*4403 | 5.12% | 0.332 | 2.163 | |
| | | | | | | LSEITKQEK | A*1101 | 7.07% | 0.191 | 1.241 | |
| | | | | | | PLSEITKQEK | A*1101 | 7.07% | 0.148 | 0.961 | |
| | | Cervix | 7.98 | 5.04% | 0.40 | SEITKQEKD | B*4403 | 5.12% | 0.112 | 0.045 | |
| | | | | | | SEITKQEKDF | B*4402 | 2.85% | 0.374 | 0.150 | |
| | | | | | | | B*4403 | 5.12% | 0.332 | 0.134 | |
| | | | | | | LSEITKQEK | A*1101 | 7.07% | 0.191 | 0.077 | |
| | | | | | | PLSEITKQEK | A*1101 | 7.07% | 0.148 | 0.059 | |
| | | Urinary tract | 20.27 | 6.79% | 1.38 | SEITKQEKD | B*4403 | 5.12% | 0.112 | 0.155 | |
| | | | | | | SEITKQEKDF | B*4402 | 2.85% | 0.374 | 0.515 | |
| | | | | | | | B*4403 | 5.12% | 0.332 | 0.457 | |
| | | | | | | LSEITKQEK | A*1101 | 7.07% | 0.191 | 0.262 | |
| | | | | | | PLSEITKQEK | A*1101 | 7.07% | 0.148 | 0.203 | |
| ABL1 | T315I | Hematopoietic and lymphoid tissue | 46.19 | 5.49% | 2.54 | FYIIIEFMTY | A*2402 | 9.29% | 0.221 | 0.561 | 9.80 |
| | | | | | | IIIEFMTYG | A*0201 | 13.66% | 0.329 | 0.835 | |
| | | | | | | YIIIEFMTY | B*1501 | 2.64% | 0.257 | 0.652 | |
| | | | | | | | B*3501 | 5.82% | 0.148 | 0.374 | |
| | | | | | | IEFMTYGNL | B*4001 | 3.26% | 0.780 | 1.979 | |
| | | | | | | | B*4403 | 5.12% | 0.121 | 0.307 | |
| | | | | | | REPPFYIII | A*2402 | 9.29% | 0.195 | 0.495 | |
| | | | | | | | B*4001 | 3.26% | 0.202 | 0.512 | |
| | | | | | | | B*4403 | 5.12% | 0.193 | 0.490 | |

**Table 1**
(*continued*)

| Gene | Mutation[a] | Tissue | Cancer incidence[b] (per 100k) | Mutation frequency[c] | MIS[d] | Peptide | Presenting allele | HLA frequency[e] | EIS[f] | GS[g] | OS[h] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | IEFMTYGNLL | B*4402 | 2.85% | 0.412 | 1.044 | |
| | | | | | | | B*4403 | 5.12% | 0.120 | 0.305 | |
| | | | | | | IIEFMTYGNL | B*3801 | 1.03% | 0.030 | 0.077 | |
| | | | | | | TREPPFYIII | B*2705 | 1.38% | 0.042 | 0.107 | |
| | | | | | | EPPFYIIIEF | A*2402 | 9.29% | 0.371 | 0.941 | |
| | | | | | | | B*3501 | 5.82% | 0.339 | 0.859 | |
| | | | | | | | B*3801 | 1.03% | 0.028 | 0.071 | |
| | | | | | | | B*4402 | 2.85% | 0.073 | 0.185 | |
| CTNNB1 | T41A | Soft tissue | 3.03 | 15.02% | 0.46 | IHSGATATA | A*0203 | 0.90% | 0.039 | 0.018 | 8.36 |
| | | | | | | | B*3801 | 1.03% | 0.084 | 0.038 | |
| | | | | | | TATAPSLSGK | A*1101 | 7.07% | 0.326 | 0.148 | |
| | | | | | | | A*6801 | 3.50% | 0.236 | 0.108 | |
| | | | | | | ATAPSLSGK | A*1101 | 7.07% | 3.535 | 1.609 | |
| | | | | | | | A*3101 | 2.40% | 0.069 | 0.031 | |
| | | | | | | | A*6801 | 3.50% | 0.493 | 0.224 | |
| | | | | | | GATATAPSL | B*5101 | 4.29% | 0.143 | 0.065 | |
| | | | | | | GIHSGATATA | A*0203 | 0.90% | 0.066 | 0.030 | |
| | | | | | | ATATAPSLSG | A*1101 | 7.07% | 0.213 | 0.097 | |
| | S45del | Kidney | 14.12 | 5.15% | 0.73 | SGATTTAPL | B*0702 | 6.71% | 0.257 | 0.187 | |
| | | | | | | ATTTAPLSGK | A*1101 | 7.07% | 0.935 | 0.680 | |
| | | | | | | | A*6801 | 3.50% | 0.200 | 0.146 | |
| | | | | | | TTTAPLSGK | A*1101 | 7.07% | 2.621 | 1.906 | |
| | | | | | | | A*6801 | 3.50% | 0.690 | 0.502 | |
| | | | | | | APLSGKGNPE | B*0702 | 6.71% | 0.239 | 0.174 | |
| | | | | | | HSGATTTAPL | B*0702 | 6.71% | 0.276 | 0.201 | |
| | S45F | Soft tissue | 3.03 | 9.86% | 0.30 | GATTTAPFL | B*5101 | 4.29% | 0.143 | 0.043 | |
| | | | | | | TTAPFLSGK | A*1101 | 7.07% | 3.535 | 1.056 | |
| | | | | | | | A*3101 | 2.40% | 0.100 | 0.030 | |
| | | | | | | | A*6801 | 3.50% | 0.965 | 0.288 | |
| | | | | | | TTTAPFLSGK | A*1101 | 7.07% | 0.713 | 0.213 | |
| | | | | | | | A*6801 | 3.50% | 0.466 | 0.139 | |
| | | | | | | TTTAPFLSG | A*1101 | 7.07% | 0.183 | 0.055 | |
| | | | | | | HSGATTTAPF | B*3501 | 5.82% | 0.520 | 0.155 | |
| | S37A | Small intestine | 1.92 | 7.77% | 0.15 | SYLDSGIHAG | A*2402 | 9.29% | 0.398 | 0.059 | |
| | | | | | | SYLDSGIHA | A*2402 | 9.29% | 0.422 | 0.063 | |
| | | | | | | YLDSGIHAGA | A*0203 | 0.90% | 0.115 | 0.017 | |
| | | | | | | GIHAGATTTA | A*0203 | 0.90% | 0.038 | 0.006 | |
| | | | | | | AGATTTAPSL | B*0702 | 6.71% | 0.362 | 0.054 | |
| | | | | | | IHAGATTTA | A*0203 | 0.90% | 0.039 | 0.006 | |
| | | | | | | | B*3801 | 1.03% | 0.084 | 0.013 | |
| KIT | D816V | Hematopoietic and lymphoid tissue | 46.19 | 9.59% | 4.43 | RVIKNDSNY | B*1501 | 2.64% | 0.372 | 1.648 | 8.35 |
| | | | | | | KICDFGLARV | A*0201 | 13.66% | 1.392 | 6.165 | |
| | | | | | | | A*3101 | 2.40% | 0.058 | 0.257 | |
| | | | | | | ARVIKNDSN | B*2705 | 1.38% | 0.062 | 0.275 | |
| HNF1A | P291fs*51 | Large intestine | 45.94 | 10.94% | 5.03 | PPQGQARDL | B*0702 | 6.71% | 1.072 | 5.386 | 7.97 |
| | | | | | | GPPQGQARDL | B*5101 | 4.29% | 0.514 | 2.581 | |
| JAK2 | V617F | Hematopoietic and lymphoid tissue | 46.19 | 45.12% | 20.84 | CFCGDENIL | A*2402 | 9.29% | 0.189 | 3.936 | 5.96 |
| | | | | | | LVLNYGVCF | B*1501 | 2.64% | 0.097 | 2.022 | |
| BRAF | V600E | Biliary tract | 1.14 | 10.06% | 0.11 | GDFGLATEK | A*1101 | 7.07% | 0.286 | 0.033 | 5.16 |
| | | Eye | 0.77 | 11.35% | 0.09 | GDFGLATEK | A*1101 | 7.07% | 0.286 | 0.025 | |
| | | Large intestine | 45.94 | 12.78% | 5.87 | GDFGLATEK | A*1101 | 7.07% | 0.286 | 1.681 | |
| | | Ovary | 12.62 | 8.70% | 1.10 | GDFGLATEK | A*1101 | 7.07% | 0.286 | 0.314 | |
| | | Skin | 19.79 | 33.51% | 6.63 | GDFGLATEK | A*1101 | 7.07% | 0.286 | 1.899 | |
| | | Thyroid | 10.60 | 39.71% | 4.21 | GDFGLATEK | A*1101 | 7.07% | 0.286 | 1.205 | |
| IDH1 | R132H | Central nervous system | 6.09 | 30.77% | 1.87 | KPIIIGHHA | B*0702 | 6.71% | 1.025 | 1.921 | 4.24 |
| | | | | | | KPIIIGHHAY | B*3501 | 5.82% | 0.903 | 1.693 | |
| | | | | | | PIIIGHHAY | B*1501 | 2.64% | 0.335 | 0.628 | |
| NRAS | Q61R | Skin | 19.79 | 7.12% | 1.41 | LLDILDTAGR | A*6801 | 3.50% | 0.527 | 0.743 | 2.72 |
| | | | | | | GREEYSAMR | B*2705 | 1.38% | 0.448 | 0.631 | |
| NRAS | Q61K | Skin | 19.79 | 6.84% | 1.35 | LLDILDTAGK | A*1101 | 7.07% | 0.996 | 1.348 | |
| RET | M918T | Thyroid | 10.60 | 28.38% | 3.01 | VKWTAIESL | B*3902 | 0.11% | 0.042 | 0.126 | 2.51 |
| | | | | | | VKWTAIESLF | A*2402 | 9.29% | 0.403 | 1.213 | |
| | | | | | | GRIPVKWTA | B*2705 | 1.38% | 0.036 | 0.108 | |
| | | | | | | GRIPVKWTAI | B*2705 | 1.38% | 0.050 | 0.151 | |
| | | | | | | KWTAIESLF | A*2402 | 9.29% | 0.304 | 0.915 | |
| PDGFRA | D842V | Soft tissue | 3.03 | 7.00% | 0.21 | ARVIMHDSN | B*2705 | 1.38% | 0.060 | 0.013 | 0.65 |
| | | | | | | RVIMHDSNY | A*1101 | 7.07% | 0.144 | 0.031 | |
| | | | | | | | B*1501 | 2.64% | 0.575 | 0.122 | |
| | | | | | | KICDFGLARV | A*0201 | 13.66% | 1.212 | 0.257 | |
| | | | | | | VIMHDSNYV | A*0201 | 13.66% | 0.500 | 0.106 | |
| | | | | | | RVIMHDSNYV | A*0201 | 13.66% | 0.553 | 0.117 | |
| MET | Y1253D | Upper aerodigestive tract | 13.51 | 8.81% | 1.19 | DMYDKEYDSV | A*0201 | 13.66% | 0.297 | 0.353 | 0.58 |
| | | | | | | KEYDSVHNK | A*1101 | 7.07% | 0.156 | 0.186 | |
| | | | | | | | B*2705 | 1.38% | 0.032 | 0.039 | |

**Table 1**
(*continued*)

| Gene | Mutation[a] | Tissue | Cancer incidence[b] (per 100k) | Mutation frequency[c] | MIS[d] | Peptide | Presenting allele | HLA frequency[e] | EIS[f] | GS[g] | OS[h] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| APC | T1556fs*3 | Pancreas | 11.70 | 6.54% | 0.77 | QEKEAEKNY | B*4403 | 5.12% | 0.075 | 0.058 | 0.14 |
| | | Small intestine | 1.92 | 5.16% | 0.10 | QEKEAEKNY | B*4403 | 5.12% | 0.075 | 0.007 | |
| | | Stomach | 7.48 | 5.84% | 0.44 | QEKEAEKNY | B*4403 | 5.12% | 0.075 | 0.033 | |
| | S1341R | Soft tissue | 3.03 | 5.25% | 0.16 | SRLSSESAR | B*2705 | 1.38% | 0.030 | 0.005 | |
| | | | | | | SRLQGSRLS | B*2705 | 1.38% | 0.012 | 0.002 | |
| | | | | | | SSRLQGSRL | B*0702 | 6.71% | 0.045 | 0.007 | |
| | | | | | | SRLSSESARH | B*2705 | 1.38% | 0.014 | 0.002 | |
| | | | | | | RLSSESARHK | A*1101 | 7.07% | 0.049 | 0.008 | |
| | | | | | | KSSRLQGSR | A*1101 | 7.07% | 0.051 | 0.008 | |
| | | | | | | SRLQGSRLSS | B*2705 | 1.38% | 0.009 | 0.001 | |
| | | | | | | KSSRLQGSRL | B*0702 | 6.71% | 0.053 | 0.008 | |

[a]HGVS notations for describing mutations is a system produced by the Human Genome Variation Society (http://www.hgvs.org/).
[b]Cancer incidence for US population from http://seer.cancer.gov.
[c]Mutation frequency calculated from the Catalogue of Somatic Mutations in Cancer (COSMIC; Forbes *et al.* 2008).
[d]MIS (Mutation Impact Score) = Frequency of mutation in given cancer × cancer incidence.
[e]Average HLA allele frequency calculated from www.allelefrequencies.net for all US ethnic groups with $n >= 500$.
[f]EIS (Epitope Impact Score) = Epitope prediction score × average frequency of presenting HLA allele.
[g]GS (Global Score) = MIS × EIS.
[h]OS (Overall Score) = Σ GS.

score is beneficial, because epitopes from distributed mutations are likely to be inferior to those derived from hotspot mutations. This is because unlike hotspot mutations, which are typically early occurring, gain of function mutations essential for tumor viability, distributed mutations are usually loss of function mutations that arise later in tumorigenesis. Loss of function mutations are undesirable immunologic targets because genes mutated in this manner become nonessential to the cell, and immune evasion can result from downregulation of mutant epitope presentation. Mutations incurred late in the tumorigenic process are generally undesirable immunologic targets for the added reason that well-developed tumors are immunoresistant, and the opportunity for prevention has passed.

To be useful as a vaccine epitope, a mutation must be immunogenic. That is, it must be contained within a peptide that is presented at the cell surface by HLA class I molecules and, ultimately, be sufficiently distinct from the wild-type version of the same peptide to cause a T-cell response. Here, we used computational epitope prediction methods to identify common tumor mutations that may also be immunogenic. It is important to note that this approach predicts only peptide-HLA binding, which is taken as a correlate of immunogenicity. Whether a given peptide is truly immunogenic also depends on it being appropriately processed and loaded on HLA molecules, and being sufficiently distinct from the presumably tolerated, nonmutated version. In the present study, we estimate the proportion of tumor mutations that could be immunogenic, and should be prioritized for further investigation. At present, the cost and throughput of laboratory immunoassays are prohibitive for screening large numbers of candidate mutations and a computational approach that shortlists candidates is warranted. For *in silico* epitope prediction tools, false-positive predictions are known to occur [39,41]. To curb this problem, only peptides predicted by two or more tools and scoring within the top ranking 5% are considered here.

Taking advantage of three well-established *in silico* epitope prediction programs, we queried in bulk every mutation or frameshift-peptide from candidate cancer genes identified, as described above, using the COSMIC database. Again, only peptides ranking within the top 5% of all possible peptide degradation products from a given gene, and predicted independently by at least two of the three programs were retained for analysis. These predictions, in conjunction with HLA class I allele frequencies for the US population (http://www.allelefrequencies.org) were used to produce an EIS for each mutant peptide (see Methods).

We screened all possible tiled peptides derived from the 20 genes and 36 mutations identified against all HLA class I variants represented by the three *in silico* epitope prediction programs, for 54,432 individual queries. For each gene, and 35 of the 36 mutations, we obtained one or more epitopes that passed filtering. There were 229 total and 159 unique peptides predicted to be presented by up to four distinct HLA class I alleles. Interestingly, of the 10 peptides with the highest EIS, seven are predicted to arise from the KRAS oncogene. Six candidates are position G12 mutants (two G12V, two G12C, one G12R, and one G12D) and one from G13D. Most KRAS mutational peptides were predicted to be high-ranking HLA binders by some of the most frequently represented alleles in the US population, which is an important factor in evaluating their EIS.

Finally, we ranked the 20 cancer genes according to an OS, which took into account all of our criteria, including the frequency of mutations within each gene in specific tumors, the prevalence of those tumors, the likelihood of mutations to be presented as HLA class I–restricted epitopes, and the population frequency of HLA alleles. The top-ranking gene was KRAS (OS =506.52), followed by NPM1 (OS =87.49) and HRAS (OS =49.24). The lowest ranked of the 20 genes was APC, with an OS of just 0.14. KRAS figures prominently because there are frequent hotspot mutations in numerous, prevalent cancers, and mutant peptides are predicted to be presented by several common HLA alleles (Table 1). Relative to KRAS, APC mutations occur infrequently in rare cancers, and are predicted to be presented weakly by rare HLA alleles. Our data illustrate other interesting features of predicted tumor epitopes. For example, we observe that a single peptide, KICDFGLARV, is shared by PDGFRA D842V and KIT D816V and, therefore, constitutes an attractive candidate epitope for simultaneous targeting of hematopoietic/lymphoid (KIT) and soft-tissue tumors (PDGFRA).

## 4. Discussion

The predicted epitopes presented here are almost certainly an underestimate of those that actually occur in cancer. There are several reasons for this. First, we have applied stringent filters for mutation frequency and epitope prediction scores that may have excluded many real epitopes. Second, our results are limited by the fact that epitope prediction tools support binding models for only a fraction of several hundred known HLA class I coding variants. For instance, BIMAS, SYFPEITHI, and IEDB support 18, 22, and 41 HLA class I coding variants, respectively. Only nine HLA class I coding variants are shared by at least two of the three epitope prediction tools and only four variants are shared by all three. Considering the

19 alleles supported by at least two programs, and frequencies of these alleles in the US population, we calculate that approximately 55% of the total US population would have at least one of these alleles, and our study is essentially blind to the remaining 45%. This is an inherent limitation of existing epitope prediction tools, and affects all epitope prediction studies. Finally, it is important to consider that mutation detection strategies have, to date, relied heavily on capillary sequencing of polymerase chain reaction–amplified coding sequence. This is a relatively insensitive approach where mutations can be obscured by coamplification of wild-type sequence from stromal cells. Single-molecule sequencing methods produce a digital readout of allele sequences and promise greater sensitivity [42]. Thus, with time, the number of candidate tumor mutational epitopes will increase. Mutations now thought to be rare may be shown to be present at more significant frequencies and new mutations are likely to be discovered by ongoing cancer genome screening efforts.

While the concept of prophylactic cancer vaccination has been discussed for some time, there have been no previous attempts to estimate the potential utility of such vaccines, in terms of the proportion of individuals that may benefit. As discussed above, the key factors to consider are tumor incidence, mutation frequency, and HLA restriction. HLA restriction rarely receives adequate attention, but is a critical issue in vaccine development, given the extensive allelic variation at this locus and the fact that efficient presentation of a given epitope is limited to a subset of HLA alleles. Here, for demonstration, we have estimated the proportion of the US population that could benefit from hypothetical multivalent vaccines targeting the relatively common colorectal and hematological/lymphoid cancers. Specifically, we determined the proportion of the US population that carries a presenting HLA allele for at least one of the mutant peptides associated with the tumor in question (Table 1). By this approach we estimate that hypothetically, 16.8% of those who would otherwise develop a hematopoietic or lymphoid cancer could be protected by vaccination with mutational epitopes predicted here, and of those who would otherwise develop colorectal cancer, 11.5% could be protected. Interestingly, given the prevalence of KRAS mutations in both pancreatic and colorectal cancer, a multivalent vaccine targeting colorectal cancer could also protect 12% of individuals who would otherwise develop pancreatic cancer.

The results presented here are purely for demonstration, but suggest meaningful levels of cancer prophylaxis could be achievable by vaccination. The results are limited by the fact that that only about half of the US population is estimated to carry at least one of the alleles considered by the epitope prediction tools employed. This taken with our incomplete knowledge of recurrent tumor mutations means that we are likely underestimating significantly the population reach of hypothetical vaccine constructs. Our projections are also based on the premise that predicted epitopes are immunogenic, and that immunogenic responses would be protective, both of which require experimental validation. There is little certainty as to which predicted epitopes correspond to high-value early driver mutations, and some may be from later onset mutations that occur in established tumors, and therefore have limited utility for immunoprevention. Finally it is, of course, not possible to predict who will or will not develop a spontaneous tumor; thus, our estimates are based on vaccination of the entire population. In principle, one could restrict vaccination to individuals carrying HLA alleles known to present epitopes contained in a vaccine construct, but the extra efforts required for HLA typing would detract from the utility of population cancer control. A more feasible strategy for targeted vaccination is in association with a cancer screening program, for example, colonoscopy for detection of colorectal tumors. Individuals with benign adenomatous polyps, identified by colonoscopy, are at risk of progression to adenocarcinoma. Progression from adenoma to adenocarcinoma is typically mediated by mutations in KRAS, BRAF, and other oncogenes, as well as loss of function mutations in TP53 and other tumor suppressors [43]. Thus, a progression blocking vaccine based on KRAS and related epitopes could be an effective preventative strategy for this tumor site.

## Acknowledgments

The mutation data was obtained from the Sanger Institute Catalogue of Somatic Mutations in Cancer ftp site, ftp://ftp.sanger.ac.uk/pub/CGP/cosmic/. This work was funded by Genome Canada, Genome British Columbia, and the Canadian Institutes of Health Research. R.A.H. is a Michael Smith Foundation for Health Research Scholar.

## References

[1] Ehrlich P. Experimentelle Karzinom-studien on Mausen. Arch Inst Exp Ther 1906;1:65.
[2] Bashford E, Murray J, Haaland M, Bowen W. General results of propagation of malignant new growths. In Bashford E, ed. Third Scientific Report on the Investigations of the Imperial Cancer Research Fund. London: Taylor and Francis; 1908.
[3] Prehn RT, Main JM. Immunity to methylcholanthrene-induced sarcomas. J Natl Cancer Inst 1957;18:769–78.
[4] Burnet FM. The concept of immunological surveillance. Prog Exp Tumor Res 1970;13:1–27.
[5] Shankaran V, Ikeda H, Bruce AT, White JM, Swanson PE, Old LJ, et al. IFNgamma and lymphocytes prevent primary tumour development and shape tumour immunogenicity. Nature 2001;410:1107–11.
[6] Koebel CM, Vermi W, Swann JB, Zerafa N, Rodig SJ, Old LJ, et al. Adaptive immunity maintains occult cancer in an equilibrium state. Nature 2007;450:903–7.
[7] Dunn GP, Old LJ, Schreiber RD. The three Es of cancer immunoediting. Annu Rev Immunol 2004;22:329–60.
[8] Willimsky G, Blankenstein T. Sporadic immunogenic tumours avoid destruction by inducing T-cell tolerance. Nature 2005;437:141–46.
[9] Cefai D, Morrison BW, Sckell A, Favre L, Balli M, Leunig M, et al. Targeting HER-2/neu for active-specific immunotherapy in a mouse model of spontaneous breast cancer. Int J Cancer 1999;83:393–400.
[10] Dakappagari NK, Douglas DB, Triozzi PL, Stevens VC, Kaumaya PT. Prevention of mammary tumors with a chimeric HER-2 B-cell epitope peptide vaccine. Cancer Res 2000;60:3782–89.
[11] Nanni P, Nicoletti G, De Giovanni C, Landuzzi L, Di Carlo E, Cavallo F, et al. Combined allogeneic tumor cell vaccination and systemic interleukin 12 prevents mammary carcinogenesis in HER-2/neu transgenic mice. J Exp Med 2001;194:1195–205.
[12] Park JM, Terabe M, Sakai Y, Munasinghe J, Forni G, Morris JC, et al. Early role of $CD4^+$ Th1 cells and antibodies in HER-2 adenovirus vaccine protection against autochthonous mammary carcinomas. J Immunol 2005;174:4228–36.
[13] Villa LL, Costa RL, Petta CA, Andrade RP, Paavonen J, Iversen OE, et al. High sustained efficacy of a prophylactic quadrivalent human papillomavirus types 6/11/16/18 L1 virus-like particle vaccine through 5 years of follow-up. Br J Cancer 2006;95:1459–66.
[14] Penn I. Malignant melanoma in organ allograft recipients. Transplantation 1996;61:274–78.
[15] Suranyi MG, Hogan PG, Falk MC, Axelsen RA, Rigby R, Hawley C, et al. Advanced donor-origin melanoma in a renal transplant recipient: Immunotherapy, cure, and retransplantation. Transplantation 1998;66:655–61.
[16] MacKie RM, Reid R, Junor B. Fatal melanoma transferred in a donated kidney 16 years after melanoma surgery. N Engl J Med 2003;348:567–68.
[17] Birkeland SA, Storm HH, Lamm LU, Barlow L, Blohme I, Forsberg B, et al. Cancer risk after renal transplantation in the Nordic countries, 1964–1986. Int J Cancer 1995;60:183–89.
[18] Mihm MC, Clemente CG, Cascinelli N. Tumor infiltrating lymphocytes in lymph node melanoma metastases: A histopathologic prognostic indicator and an expression of local immune response. Lab Invest 1996;74:43–7.
[19] Clemente CG, Mihm MC, Bufalino R, Zurrida S, Collini P, Cascinelli N. Prognostic value of tumor infiltrating lymphocytes in the vertical growth phase of primary cutaneous melanoma. Cancer 1996;77:1303–10.
[20] Zhang L, Conejo-Garcia JR, Katsaros D, Gimotty PA, Massobrio M, Regnani G, et al. Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer. N Engl J Med 2003;348:203–13.
[21] Reuschenbach M, von Knebel Doeberitz M, Wentzensen N. A systematic review of humoral immune responses against tumor antigens. Cancer Immunol Immunother 2009;58:1535–44.
[22] Wang HY, Zhou J, Zhu K, Riker AI, Marincola FM, Wang RF. Identification of a mutated fibronectin as a tumor antigen recognized by CD4+ T cells: Its role in extracellular matrix formation and tumor metastasis. J Exp Med 2002;195:1397–406.
[23] Gaudin C, Kremer F, Angevin E, Scott V, Triebel F. A hsp70-2 mutation recognized by CTL on a human renal cell carcinoma. J Immunol 1999;162:1730–8.

[24] Huang J, El-Gamil M, Dudley ME, Li YF, Rosenberg SA, Robbins PF. T cells associated with tumor regression recognize frameshifted products of the CDKN2A tumor suppressor gene locus and a mutated HLA class I gene product. J Immunol 2004;172:6057–64.
[25] Linard B, Bezieau S, Benlalam H, Labarriere N, Guilloux Y, Diez E, et al. A ras-mutated peptide targeted by CTL infiltrating a human melanoma lesion. J Immunol 2002;168:4802–8.
[26] Robbins PF, El-Gamil M, Li YF, Kawakami Y, Loftus D, Appella E, et al. A mutated beta-catenin gene encodes a melanoma-specific antigen recognized by tumor infiltrating lymphocytes. J Exp Med 1996;183:1185–92.
[27] Kawakami Y, Wang X, Shofuda T, Sumimoto H, Tupesis J, Fitzgerald E, et al. Isolation of a new melanoma antigen, MART-2, containing a mutated epitope recognized by autologous tumor-infiltrating T lymphocytes. J Immunol 2001; 166:2871–7.
[28] Lollini PL, Cavallo F, Nanni P, Forni G. Vaccines for tumour prevention. Nat Rev Cancer 2006;6:204–16.
[29] Gjertsen MK, Bjorheim J, Saeterdal I, Myklebust J, Gaudernack G. Cytotoxic CD4+ and CD8+ T lymphocytes, generated by mutant p21-ras (12Val) peptide vaccination of a patient, recognize 12Val-dependent nested epitopes present within the vaccine peptide and kill autologous tumour cells carrying this mutation. Int J Cancer 1997;72:784–90.
[30] Ito D, Visus C, Hoffmann TK, Balz V, Bier H, Appella E, et al. Immunological characterization of missense mutations occurring within cytotoxic T cell-defined p53 epitopes in HLA-A*0201+ squamous cell carcinomas of the head and neck. Int J Cancer 2007;120:2618–24.
[31] Messaoudi I, Guevara Patino JA, Dyall R, LeMaoult J, Nikolich-Zugich J. Direct link between *mhc* polymorphism, T cell avidity, and diversity in immune defense. Science 2002;298:1797–800.
[32] Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. Br J Cancer 2004;91:355–58.
[33] Forbes S, Clements J, Dawson E, Bamford S, Webb T, Dogan A, et al. COSMIC 2005. Br J Cancer 2006;94:318–22.
[34] Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, et al. The Catalogue of Somatic Mutations in Cancer (COSMIC). Curr Protoc Hum Genet 2008;Chapter 10:Unit 10.11.
[35] Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. SYFPEITHI: Database for MHC ligands and peptide motifs. Immunogenetics 1999;50:213–19.
[36] Parker KC, Bednarek MA, Coligan JE. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J Immunol 1994;152:163–75.
[37] Peters B, Sidney J, Bourne P, Bui HH, Buus S, Doh G, et al. The design and implementation of the immune epitope database and analysis resource. Immunogenetics 2005;57:326–36.
[38] Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BMC Bioinformatics 2005;6:132.
[39] Kessler JH, Melief CJ. Identification of T-cell epitopes for cancer immunotherapy. Leukemia 2007;21:1859–74.
[40] Gulukota K, DeLisi C. HLA allele selection for designing peptide vaccines. Genet Anal 1996;13:81–6.
[41] Kessler JH, Mommaas B, Mutis T, Huijbers I, Vissers D, Benckhuijsen WE. Competition-based cellular peptide binding assays for 13 prevalent HLA class I alleles using fluorescein-labeled synthetic peptides. Hum Immunol 2003;64: 245–55.
[42] Pushkarev D, Neff NF, Quake SR. Single-molecule sequencing of an individual human genome. Nat Biotechnol 2009;27:847–52.
[43] Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. Genetic alterations during colorectal-tumor development. N Engl J Med 1988; 319:525–32.